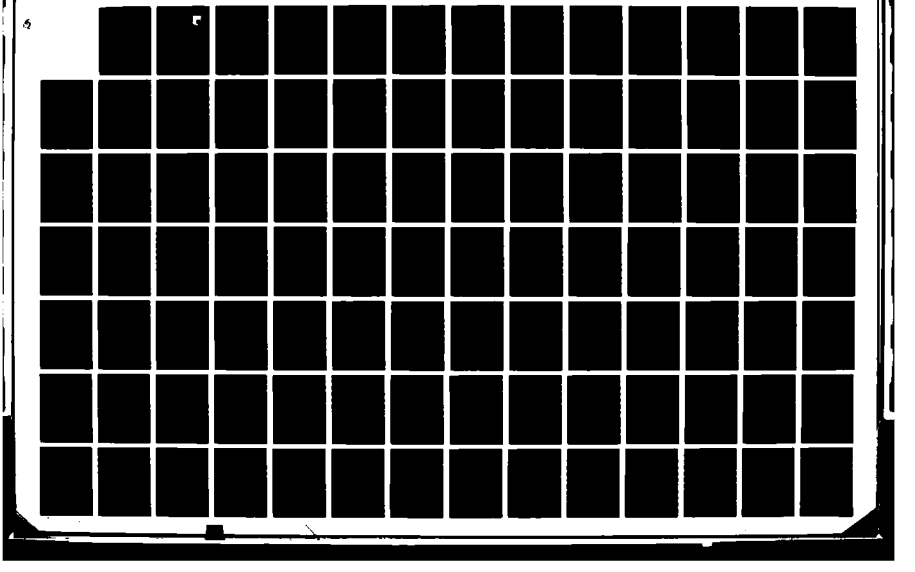
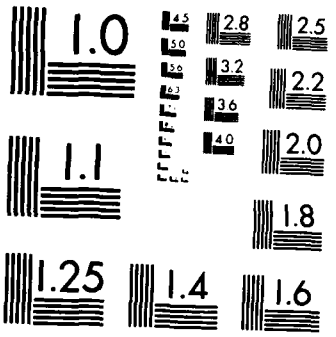
AD-A188 378 A NEURONAL MODEL OF CLASSICAL CONDITIONING(U) AIR FORCE 1/2
WRIGHT AERONAUTICAL LABS WRIGHT-PATTERSON AFB OH
A H KLOPF OCT 87 AFWAL-TR-87-1139

UNCLASSIFIED F/G 12/9 NL





AFWAL-TR-87-1139

A NEURONAL MODEL OF CLASSICAL CONDITIONING

A. Harry Klopf
Information Processing Technology Branch
System Avionics Division

October 1987

INTERIM REPORT for period October 1979 - September 1987

AVIONICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6543

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

A. HARRY KLOPF
Principal Investigator
Advanced Systems Research Group
Avionics Laboratory

RICHARD C. JONES, Chief
Advanced Systems Research Group
Avionics Laboratory

FOR THE COMMANDER

CHARLES H. KRUEGER, JR
Acting Chief
System Avionics Division
Avionics Laboratory

If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFWAL/AAAT, WPAFB, OH 45433-6543 to help us maintain a current mailing list.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFWAL-TR-87-1139	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Avionics Laboratory, AF Wright Aeronautical Labs 6b. OFFICE SYMBOL (If applicable) AFWAL/AAAT-3	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) AFWAL/AAAT-3 Wright-Patterson AFB OH 45433	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332	10. SOURCE OF FUNDING NOS PROGRAM ELEMENT NO: 61102F; PROJECT NO: 2312; TASK NO: R1; WORK UNIT NO: 02
11. TITLE (Include Security Classification) A Neuronal Model of Classical Conditioning (U)	
12. PERSONAL AUTHOR(S) A. Harry Klopf	
13a. TYPE OF REPORT Interim 13b. TIME COVERED FROM 10/79 TO 9/87	14. DATE OF REPORT (Yr., Mo., Day) October 1987 15. PAGE COUNT 159
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB. GR.
12	09	
05	08	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Neuronal Model; Learning; Neural Network Model; Adaptive Network; Artificial Intelligence; Classical Conditioning

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A neuronal model of classical conditioning is proposed. The model is most easily described by contrasting it with a still influential neuronal model first analyzed by Hebb (1949). It is proposed that the Hebbian model be modified in three ways to yield a model more in accordance with animal learning phenomena. ~~First~~, instead of correlating pre- and postsynaptic levels of activity, changes in pre- and postsynaptic levels of activity should be correlated to determine the changes in synaptic efficacy that represent learning. Second, instead of correlating approximately simultaneous pre- and postsynaptic signals, earlier changes in presynaptic signals should be correlated with later changes in postsynaptic signals. Third, a change in the efficacy of a synapse should be proportional to the current efficacy of the synapse, accounting for the initial positive acceleration in the s-shaped acquisition curves observed in animal learning. The resulting model, termed a drive-reinforcement model of single neuron function, suggests that nervous system activity can be understood in terms of two

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL A. Harry Klopf	22b. TELEPHONE NUMBER (Include Area Code) (513) 255-7649 22c. OFFICE SYMBOL AFWAL/AAAT-3

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

19. Abstract (Continued)

classes of neuronal signals: Drives that are defined to be signal levels and reinforcers that are defined to be changes in signal levels. Defining drives and reinforcers in this way, in conjunction with the neuronal model, suggests a basis for a neurobiological theory of learning. The proposed neuronal model is an extension of the Sutton-Barto (1981) model which, in turn, can be seen as a temporally refined extension of the Rescorla-Wagner (1972) model. It is shown that the proposed neuronal model predicts the basic categories of classical conditioning phenomena including delay and trace conditioning, conditioned and unconditioned stimulus duration and amplitude effects, partial reinforcement effects, interstimulus interval effects including simultaneous conditioning, second-order conditioning, conditioned inhibition, extinction, reacquisition effects, backward conditioning, blocking, overshadowing, compound conditioning, and discriminative stimulus effects. The neuronal model also eliminates some inconsistencies with the experimental evidence that occur with the Rescorla-Wagner and Sutton-Barto models. Implications of the neuronal model for animal learning theory, connectionist and neural network modeling, artificial intelligence, adaptive control theory, and adaptive signal processing are discussed. It is concluded that real-time learning mechanisms that do not require evaluative feedback from the environment are fundamental to natural intelligence and may have implications for artificial intelligence. Experimental tests of the model are suggested.

ACKNOWLEDGEMENTS

This research was supported by the Life Sciences Directorate of the Air Force Office of Scientific Research under Task 2312 R1.

Jim Morgan has contributed substantially throughout the course of this research. In particular, I want to acknowledge many valuable discussions and the software Jim wrote for the computer simulations of the neuronal models.

I am grateful to the following people for comments on an earlier draft of this report: Andy Barto, Diana Blazis, Jack Byrne, John Desmond, Bruce Edson, Chuck Hendrix, Joan Klopf, John Moore, Jim Morgan, Libby Patterson, Rick Ricart, and Rich Sutton.

TABLE OF CONTENTS

SECTION		PAGE
1.	INTRODUCTION	1
2.	THE NEURONAL MODEL	3
	Qualitative Description.	3
	Mathematical Specification	6
	Properties of the Model.	12
	Refinement of the Model.	13
	Derivation and Evolution of the Drive-Reinforcement Model from Earlier Models.	16
3.	CLASSICAL CONDITIONING: PREDICTIONS OF THE NEURONAL MODEL	24
	Delay Conditioning	30
	CS and US Duration Effects	33
	CS and US Amplitude Effects.	38
	CS Preexposure Effects	40
	Partial Reinforcement Effects.	40
	Trace Conditioning	42
	Interstimulus Interval Effects Including Simultaneous Conditioning.	45
	Second-Order Conditioning.	47
	Conditioned Inhibition	48
	Extinction and Reacquisition Effects	53
	Backward Conditioning.	57
	Blocking and Overshadowing	59
	Compound Conditioning.	63
	Discriminative Stimulus Effects.	67

TABLE OF CONTENTS (CONCLUDED)

SECTION		PAGE
	A Variant of the Drive-Reinforcement Neuronal Model	70
	Summary	72
4.	DRIVES AND REINFORCERS	73
	Definitions	73
	Relationship of the Drive-Reinforcement Neuronal Model to Animal Learning Theory	79
	A Drive-Reinforcement Theory of Learning	86
5.	EXPERIMENTAL TESTS	90
6.	DISCUSSION	97
	Connectionist and Neural Network Modeling	97
	Artificial Intelligence	108
	Adaptive Control Theory and Adaptive Signal Processing	112
	Memory and Learning	114
7.	CONCLUDING REMARKS	116
REFERENCES		118
APPENDIX:	Parameter Specifications for the Computer Simulations of the Neuronal Models	143
	Drive-Reinforcement Model	143
	Hebbian Model	144
	Sutton-Barto Model	144

LIST OF ILLUSTRATIONS

FIGURE		PAGE
1.	A Model of a Single Neuron with n Synapses	9
2.	Examples of how the Drive-Reinforcement Learning Mechanism Alters the Onsets and Offsets of Pulse Trains for a Single Theoretical Neuron	14
3.	The Drive-Reinforcement Neuronal Model Employed in the Computer Simulations.	25
4.	Acquisition Curves in Simulated Delay Conditioning Experiments with (a) Hebbian, (b) Sutton-Barto, and (c) Drive-Reinforcement Neuronal Models.	31
5.	Effect of CS Duration in Simulated Delay Conditioning Experiments with (a) Hebbian, (b) Sutton-Barto, and (c) Drive-Reinforcement Neuronal Models.	35
6.	The Drive-Reinforcement Model's Predictions of the Effects of US Duration.	37
7.	The Drive-Reinforcement Model's Predictions of the Effects of CS Amplitude	39
8.	The Drive-Reinforcement Model's Predictions of the Effects of US Amplitude	41
9.	The Drive-Reinforcement Model's Predictions of the Effects of Partial Reinforcement.	43
10.	The Drive-Reinforcement Model's Predictions of the Effects of Trace Conditioning	44
11.	The Drive-Reinforcement Model's Predictions of the Effect of the Interstimulus Interval.	46
12.	The Drive-Reinforcement Model's Predictions of the Effects of Second-Order Conditioning.	49
13.	Results of a Simulated Classical Conditioning Experiment Modeled after Experiments Performed by Pavlov (1927), in which Conditioned Excitation, Conditioned Inhibition, and Extinction Paradigms are Employed.	51

LIST OF ILLUSTRATIONS (CONCLUDED)

FIGURE		PAGE
14.	Results of a Simulated Three-Stage Classical Conditioning Experiment Examining Reacquisition Effects	56
15.	Results of Simulated Classical Conditioning Experiments in which the Drive-Reinforcement Model's Predictions for (a) Forward and (b) Backward Conditioning are Compared.	58
16.	The Drive-Reinforcement Model's Predictions of the Effects of a Blocking Stimulus.	61
17.	The Drive-Reinforcement Model's Predictions of the Effects of Stimulus Salience on Compound Conditioning.	64
18.	Results of Simulated Compound Conditioning Experiments in which the Drive-Reinforcement Model's Predictions for Reinforced and Nonreinforced CSs are Compared.	66
19.	Results of Simulated Compound Conditioning Experiments in which the Drive-Reinforcement Model's Predictions of the Effects of Discriminative Stimuli were Determined for a More Complex Case than that Portrayed in Figure 18	68
20.	Results of a Simulated Blocking Experiment that was Identical to that Reported in Figure 16 Except that a Variant of the Drive-Reinforcement Model was Employed.	71

SECTION 1

INTRODUCTION

Pavlov (1927) and Hebb (1949) were among the first investigators to extensively analyze possible relationships between the behavior of whole animals and the behavior of single neurons. Building on Pavlov's experimental foundation, Hebb's theoretical analyses led him to a model of single neuron function that continues to be relevant to the theoretical and experimental issues of learning and memory. There had been earlier attempts to develop such neuronal models. Among them were the models of Freud (1895), Rashevsky (1938) and McCulloch and Pitts (1943) but, to this day, the neuronal model proposed by Hebb has remained the most influential among theorists. Current theorists who have utilized variants of the Hebbian model include Anderson, Silverman, Ritz, and Jones (1977), Kohonen (1977), Grossberg (1982), Levy and Desmond (1985), Hopfield and Tank (1986), and Rolls (1987).

In this report, I will suggest several modifications to the Hebbian neuronal model. The modifications yield a model which will be shown to be more nearly in accord with animal learning phenomena that are observed experimentally. The model to be proposed is an extension of the Sutton-Barto (1981) model.

After defining the neuronal model, first qualitatively and then mathematically, I will show, by means of computer simulations, that the neuronal model predicts the basic categories of

classical conditioning phenomena. Then, I will discuss the neuronal model in more general theoretical terms, with particular reference to the psychological notions of drives and reinforcers. My conclusion will be that the model offers a way of defining drives and reinforcers at a neuronal level such that a neurobiological basis is suggested for animal learning. In the theoretical context that the neuronal model provides, I will suggest that drives, in their most general sense, are simply signal levels in the nervous system and reinforcers, in their most general sense, are simply changes in signal levels. This seems too simple and, indeed, it is - but I hope to show that it is not that much too simple. I will attempt to make a case for drives and reinforcers being viewed, in their essence, as signal levels in the nervous system and as changes in signal levels, respectively. The result will be a theoretical framework based on what I propose to call a drive-reinforcement model of single neuron function.

SECTION 2

THE NEURONAL MODEL

Qualitative description

I will begin by defining the drive-reinforcement neuronal model in qualitative terms. It will be easiest to do this by contrasting the model with the Hebbian model. Hebb (1949) suggested that the efficacy of a plastic synapse increases whenever the synapse is active in conjunction with activity of the postsynaptic neuron. Thus, Hebb was proposing that learning (i.e., changes in the efficacy of synapses) is a function of correlations between approximately simultaneous pre- and postsynaptic levels of neuronal activity.

I wish to suggest three modifications to the Hebbian model:

(a) Instead of correlating pre- and postsynaptic levels of activity, changes in presynaptic levels of activity should be correlated with changes in postsynaptic levels of activity. In other words, instead of correlating signal levels on the input and output sides of the neuron, the first derivatives of the input and output signal levels should be correlated.

(b) Instead of correlating approximately simultaneous pre- and postsynaptic signal levels, earlier presynaptic signal levels should be correlated with later postsynaptic signal levels. More precisely and consistent with (a), earlier changes in presynaptic signal

levels should be correlated with later changes in postsynaptic signal levels. Thus, sequentiality replaces simultaneity in the model. The interval between correlated changes in pre- and postsynaptic signal levels is suggested to range up to that of the maximum effective interstimulus interval in delay conditioning.

(c) A change in the efficacy of a synapse should be proportional to the current efficacy of the synapse, accounting for the initial positive acceleration in the s-shaped acquisition curves observed in animal learning.

A refinement of the model will be noted now and discussed more fully later. The ability of the neuronal model to predict animal learning phenomena is improved if, instead of correlating positive and negative changes in neuronal inputs with changes in neuronal outputs, only positive changes in inputs are correlated with changes in outputs. To clarify this, positive changes in inputs refer to increases in the frequency of action potentials at a synapse, whether the synapse is excitatory or inhibitory. Negative changes in inputs refer to decreases in the frequency of action potentials at a synapse, whether the synapse is excitatory or inhibitory. Furthermore, the changes in frequencies of action potentials I'm referring to will be relatively abrupt, occurring within about a second or less. It is hypothesized that more gradual and long-term changes in the frequency of action potentials at a synapse do not trigger the neuronal learning mechanism.

After the neuronal model has been defined precisely and the results of computer simulations have been presented, it will be seen that this model of neuronal function bears the following relationship to models of whole animal behavior. In general, changes in presynaptic frequencies of firing will reflect the onsets and offsets of conditioned stimuli. In general, changes in postsynaptic frequencies of firing will reflect increases or decreases in levels of drives (with drives being defined more broadly than has been customary in the past). In the case of the neuronal model, changes in the levels of drives (which will usually manifest as changes in postsynaptic frequencies of firing) will be associated with reinforcement. With regard to the behavior of whole animals, the notion that changes in drive levels constitute reinforcement has been a fundamental part of animal learning theory since the time of Hull (1943) and Mowrer (1960). Here, I am taking the notion down to the level of the single neuron. Changes in signal levels, which play a fundamental role in the neuronal model being proposed, have long been recognized to be of importance. For example, Berlyne (1973, p. 16) notes that "many recent theorists have been led from different starting points to the conclusion that hedonic value is dependent above all on changes in level of stimulation or level of activity. They include McClelland, Atkinson, Clark and Lowell (1953), Premack (1959), Helson (1964), and Fowler (1971)."

Before concluding this introduction to the drive-reinforcement neuronal model, it will be useful to briefly note how the model relates to earlier models from which it derives. The derivation and evolution of the model will be discussed more fully later. As has already been

indicated, the drive-reinforcement model is an extension of the Sutton-Barto (1981) model. The Sutton-Barto model, in turn, can be viewed as a temporally refined extension of the Rescorla-Wagner (1972) model. I will show that the drive-reinforcement model eliminates some shortcomings of the Rescorla-Wagner and Sutton-Barto models. Both of the latter models predict strictly negatively accelerated acquisition or learning curves. The Rescorla-Wagner model also predicts extinction of conditioned inhibition. Consistent with the experimental evidence, it will be seen below that the drive-reinforcement model predicts (a) an acquisition curve that is initially positively accelerating and subsequently negatively accelerating and (b) conditioned inhibition that does not extinguish. In addition, the drive-reinforcement model solves some problems with conditioned stimulus duration effects that arise in the case of the Sutton-Barto model.

Mathematical specification

The proposed neuronal model may be defined precisely as follows. The input-output relationship of a neuron will be modeled in a fashion that is customary among neural network modelers. Namely, it will be assumed that single neurons are forming weighted sums of their excitatory and inhibitory inputs and then, if the sum equals or exceeds the threshold, the neuron fires. Such a model of a neuron's input-output relationship can be based on the view that neuronal signals are binary (either a neuron fires or it doesn't) or on the view that neuronal signals are real-valued (reflecting some measure of the frequency of

firing of neurons as a function of the amount by which the neuronal threshold is exceeded). Here, the latter view will be adopted. Neuronal input and output signals will be treated as frequencies. This approach to modeling neuronal input-output relationships is consistent with experimental evidence reviewed by Calvin (1975).

Mathematically, then, the neuronal input-output relationship may be specified as follows:

$$y(t)=\sum_{i=1}^{n} w_i(t)x_i(t)-\theta \qquad (1)$$

where $y(t)$ is a measure of the postsynaptic frequency of firing at discrete time t; n is the number of synapses impinging on the neuron; $w_i(t)$ is the efficacy of the i^{th} synapse; $x_i(t)$ is a measure of the frequency of action potentials at the i^{th} synapse and θ is the neuronal threshold. The synaptic efficacy, $w_i(t)$, can be positive or negative, corresponding to excitatory or inhibitory synapses, respectively. Also, $y(t)$ is bounded such that $y(t)$ is greater than or equal to zero and less than or equal to the maximal output frequency, $y'(t)$, of the neuron. Negative values of $y(t)$ have no meaning as they would correspond to negative frequencies of firing.

To complete the mathematical specification of the neuronal model, the learning mechanism described earlier in qualitative terms remains to be presented. The learning mechanism may be specified as follows:

$$\Delta w_i(t)= \Delta y(t) \sum_{j=1}^{\tau} c_j |w_i(t-j)| \Delta x_i(t-j) \qquad (2)$$

where $\Delta w_i(t)=w_i(t+1)-w_i(t)$, $\Delta y(t)=y(t)-y(t-1)$, and $\Delta x_i(t-j)=x_i(t-j)-x_i(t-j-1)$. $\Delta w_i(t)$ represents the change in the efficacy of the i^{th} synapse at time t, yielding the adjusted or new efficacy of the synapse at time $t+1$. $\Delta x_i(t-j)$ represents a

presynaptic change in signal level at time, t-j, and $\Delta y(t)$ represents the postsynaptic change in signal level at time t. τ is the longest interstimulus interval, measured in discrete time steps, over which delay conditioning is effective and c_j is an empirically established learning rate constant which is proportional to the efficacy of conditioning when the interstimulus interval is j. The remaining symbols are defined as in equation (1). A diagram of the neuron modeled by equations (1) and (2) is shown in Figure 1.

Generally, in interpreting and working with equation (2), I have adopted the following assumptions, consistent with what is known of learning involving the skeletal reflexes. I usually consider each discrete time step, t, to be equal to one-half second. This is a meaningful interval over which to obtain measures of the pre- and postsynaptic frequencies of firing, $x_i(t)$ and $y(t)$. Also, it is probably a reasonable interval of time with respect to the learning processes underlying changes in synaptic efficacy. For example, the optimal interstimulus interval for classically conditioning a skeletal reflex is nominally one-half second [optimal interstimulus intervals vary from about 200 to 500 ms depending on the species and the response system within the species (see review by Woody, 1982)], and very little or no conditioning is observed with intervals approaching zero or exceeding three seconds (Frey and Ross, 1968; McAllister, 1953; Russell, 1966; Moore and Gormezano, 1977). Thus, in equation (2), indexing starts with j equal to 1 because c_0 is equal to zero, reflecting the fact that no conditioning is observed with an interstimulus interval of zero. c_1 is assigned the maximal value reflecting the fact that one-half second is

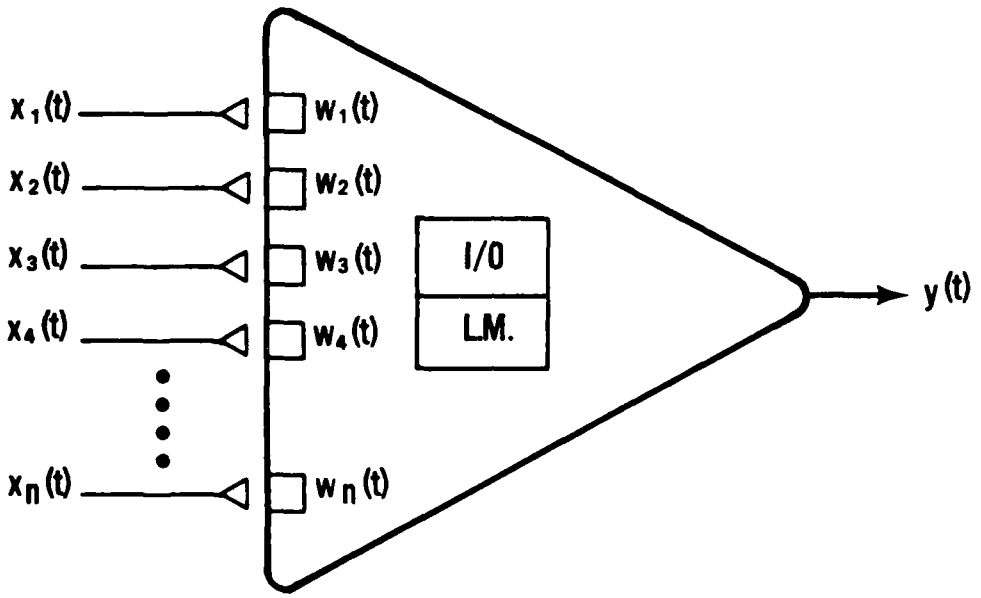


Figure 1. A model of a single neuron with n synapses. Presynaptic frequencies of firing are represented by $x_i(t)$, synaptic efficacies by $w_i(t)$, and the postsynaptic frequency of firing by $y(t)$. The input-output (I/O) relationship is specified by equation (1) and the learning mechanism (L.M.) is specified by equation (2) in the text.

(approximately) the optimal interstimulus interval. Then, c_{j+1} is less than c_j for the remaining c-values, reflecting the decreasing efficacy of conditioning as the interstimulus interval increases beyond one-half second. τ is normally set equal to 5 because, when j equals 6 (corresponding to an interstimulus interval of three seconds), little or no conditioning would occur so c_6 would be approximately equal to zero.

A lower bound is set on the absolute values of the synaptic weights, $w_i(t)$. The bound is near but not equal to zero because synaptic weights appear as factors on the right side of equation (2). It can be seen that the learning mechanism would cease to yield changes in synaptic efficacy for any synapse whose efficacy reached zero; i.e., $\Delta w_i(t)$ would henceforth always equal zero. A lower bound on the absolute values of synaptic weights results in excitatory weights always remaining excitatory (positive) and inhibitory weights always remaining inhibitory (negative); i.e., synaptic weights do not cross zero. This is consistent with the known physiology of synapses (Eccles, 1964). A nonzero lower bound on the efficacy of synapses is also consistent with evidence suggesting that potential conditioned stimuli are weakly connected to unconditioned responses prior to conditioning (Gould, 1986; Schwartz, 1978; Pavlov, 1927). Also, a nonzero lower bound on the efficacy of synapses models the notion that a synapse must have some effect on the postsynaptic neuron in order for the postsynaptic learning mechanism to be triggered. That learning mechanisms are postsynaptic, at least in phylogenetically advanced organisms, has been well argued by McNaughton, Barnes, and Rao (1984). In the case of the mammalian central nervous system, Thompson, McCormick, Lavond, Clark, Kettner, and Mauk (1983) note

that what little evidence now exists is perhaps more consistent with the hypothesis of postsynaptic rather than presynaptic learning mechanisms.

In general, it is expected that the efficacy of synapses, $w_i(t)$, is variable and under the control of the neuronal learning mechanism. However, some synapses can be expected to have fixed weights; i.e., weights that are innate and unchangeable. This may be true for many or most synapses in the autonomic nervous system. In the somatic nervous system, it is likely that many more synapses and perhaps most are variable or "plastic". In the case of the drive-reinforcement neuronal model, it will be assumed that synapses mediating conditioned stimuli have variable weights and that synapses mediating unconditioned stimuli have fixed weights. The innately specified synaptic weights that are assumed to mediate unconditioned stimuli are expected to reflect the evolutionary history of the organism.

Let us now consider what is happening in equation (2). As the specification of the learning mechanism for the drive-reinforcement neuronal model, equation (2) suggests how the efficacy of a synapse changes as a function of four factors: (1) learning rate constants, c_j, that are assumed to be innate; (2) the absolute value, $|w_i(t-j)|$, of the efficacy of the synapse at time, t-j, when the change in presynaptic level of activity occurred; (3) the change in presynaptic level of activity, $\Delta x_i(t-j)$; and (4) the change in postsynaptic level of activity, $\Delta y(t)$.

One way of visualizing either the Hebbian or the drive-reinforcement learning mechanism is in terms of a temporal window that slides along the

time line as learning occurs, changing the efficacy of synapses as it moves along. In the case of the Hebbian model, the learning mechanism employs a temporal window that is, in effect, only one time step wide. The learning mechanism slides along the time line, modifying the efficacy of synapses proportional to (1) a learning rate constant, (2) the presynaptic level of activity, and (3) the postsynaptic level of activity. (The Hebbian model will be presented in mathematical form later.) In the case of the drive-reinforcement model, the learning mechanism employs a temporal window that is $\tau+1$ time steps wide. The learning mechanism slides along the time line modifying the efficacy of synapses proportional to (1) learning rate constants, (2) the efficacy of synapses, (3) changes in presynaptic levels of activity and (4) changes in postsynaptic levels of activity. It can be seen that the Hebbian learning mechanism correlates approximately simultaneous signal levels and the drive-reinforcement learning mechanism correlates temporally separated derivatives of signal levels. (In the case of the drive-reinforcement model, I am not suggesting that a neuron would have to compute anything as refined as a first derivative. A first-order difference will suffice, as will be demonstrated later.) The differences in the behavior of the Hebbian and the drive-reinforcement learning mechanisms will be examined below when the results of computer simulations of both models are presented.

Properties of the model

The drive-reinforcement neuronal model suggests that what neurons are learning to do is to anticipate or predict the onsets and offsets of pulse trains. By pulse trains, I mean sequences or clusters of action

potentials in axons. The model neuron learns to predict the onsets and offsets of pulse trains representing unconditioned stimuli, utilizing the onsets of pulse trains representing conditioned stimuli. This will become evident when the results of computer simulations are presented. It will be seen that the learning mechanism moves the onsets and offsets of pulse trains to earlier points in time. Fundamentally, the learning mechanism is a shaper of pulse trains. The efficacy of a synapse changes in a direction such that the neuron comes to anticipate the unconditioned response; i.e., the conditioned stimulus comes to produce the conditioned response prior to the occurrence of the unconditioned stimulus and the unconditioned response. The way the drive-reinforcement neuronal learning mechanism shapes pulse trains is illustrated in Figure 2. Many investigators, including Pavlov (1927), have pointed to the anticipatory or predictive nature of conditioning phenomena [e.g., see Kamin (1968, 1969), Rescorla and Wagner (1972), Dickinson and Mackintosh (1978), and Sutton and Barto (1981)].

Refinement of the model

The drive-reinforcement neuronal learning mechanism, as defined by equation (2), can be refined in a way that improves the model's ability to predict animal learning phenomena. The refinement, as briefly noted earlier, involves allowing only positive changes in presynaptic signal levels to trigger the neuronal learning mechanism. In other words, $\Delta x_i(t-j)$ must be greater than zero. If $\Delta x_i(t-j)$ is less than zero, it is then set equal to zero for the purpose of calculating $\Delta w_i(t)$ in equation (2).

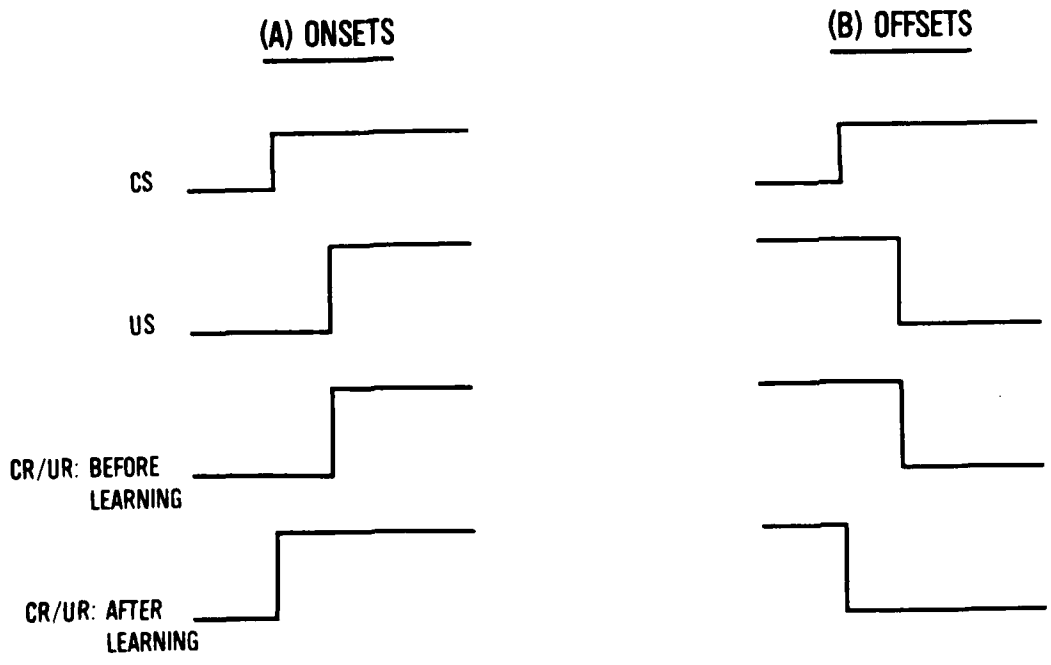


Figure 2. Examples of how the drive-reinforcement learning mechanism alters the onsets and offsets of pulse trains for a single theoretical neuron. Panels (a) and (b) show the effects of unconditioned stimulus onset and offset, respectively. In each example, the conditioned stimulus (CS) is followed by an unconditioned stimulus (US), both of which represent presynaptic signals. The two presynaptic signals are assumed to be mediated by separate synapses, with the CS-mediating synapse having a variable efficacy (weight) under the control of the neuronal learning mechanism. The conditioned and unconditioned response (CR and UR) before and after learning (i.e., before and after a number of presentations of the CS-US pair) are shown below the wave forms for the CS and US pulse trains. The conditioned and unconditioned response (CR/UR) represents the postsynaptic frequency of firing of the neuron. In panels (a) and (b), it is seen that the onset and offset of firing, respectively, occurs earlier in time after learning. Thus, in each case, the neuron has learned to anticipate the unconditioned response by learning to start firing earlier (panel a) or stop firing earlier (panel b).

There is an intuitive basis for this refinement. A negative change in presynaptic signal level means that the presynaptic signal is falling away; i.e., that it is headed toward zero. If such a negative change in presynaptic signal level were to trigger the neuronal learning mechanism and possibly cause a synaptic weight to change, then a synaptic weight would have changed for a synapse that just ceased to carry the signal that caused the change. That is to say, the relevant part of the signal on which the synaptic weight should operate would no longer be present. Some residual portion of the signal might still be present after the negative change in presynaptic signal level. However, the residual portion of the signal is not relevant because it might have been there long before the negative change in presynaptic signal level and might be there long afterward. With the drive-reinforcement neuronal learning mechanism, only the dynamic part of the signal is relevant, as will be more clearly seen after the computer simulations are presented. This is not to suggest that a drive-reinforcement learning mechanism would preclude learning about negative changes in levels of stimuli. However, if such changes are to trigger a drive-reinforcement learning mechanism, it is suggested that they would have to be, in effect, inverted, such that they would manifest in some part of the nervous system as positive changes in signal levels.

Allowing only positive changes in presynaptic signal levels to trigger the neuronal learning mechanism is part of a strategy of not changing a synaptic weight unless there is good reason to believe the weight change will be useful. Such a strategy seems reasonable because, in a neural network, there is always the possibility that a synaptic

weight change will interfere with or constitute overwriting of a previous weight change. Thus, weight changes are to be minimized.

The rationale offered above for refining the learning mechanism does not constitute a rigorous argument. However, it is hoped that the rationale provides some insight into why the refinement might make sense. Later, a more rigorous approach will be taken. It will be shown that the basic categories of classical conditioning phenomena are predicted by the neuronal model when only positive changes in presynaptic signal levels are allowed to trigger the learning mechanism. Then, it will be shown how the model's predictions deviate from the experimental evidence when both positive and negative changes in presynaptic signal levels can trigger changes in synaptic weights.

Derivation and evolution of the drive-reinforcement model from earlier models

Having defined the neuronal model in qualitative and mathematical terms, I will now describe the model's derivation and evolution from earlier neuronal models. The neuronal learning mechanisms that have been proposed, leading to the drive-reinforcement model, will be portrayed in two ways: (1) by means of the sequence of critical events that have been hypothesized to lead to learning and (2) by means of the equation that characterizes the learning mechanism. As it is customary to number equations, I will also number the critical event sequences so that I can refer to them later. To distinguish them from the equation numbers, an "S" will be added as a prefix to the critical event sequence numbers.

Hebb suggested that the sequence of critical events for learning was simple:

$$x_i(t) \rightarrow y(t) \rightarrow \Delta w_i(t) \qquad \text{(S-1)}$$

In other words, presynaptic activity, $x_i(t)$, followed directly by postsynaptic activity, $y(t)$, was hypothesized to result in a change, $\Delta w_i(t)$, in the efficacy of the associated synapse. (The convention adopted in this report is that when presynaptic activity, x_i, is a direct cause of postsynaptic activity, y, then x_i and y will have the same time step, t, associated with them.) The equation for the Hebbian learning mechanism may be written as follows:

$$\Delta w_i(t) = c x_i(t) y(t) \qquad (3)$$

where c is a learning rate constant and the other symbols are as defined earlier.

Hebb's model is an example of a simple real-time learning mechanism. Real-time learning mechanisms emphasize the temporal association of signals: each critical event in the sequence leading to learning has a time of occurrence associated with it and this time plays a fundamental role in the computations that yield changes in the efficacy of synapses. It should be noted that "real-time", in this context, does not mean continuous time as contrasted with discrete time nor does it refer to a learning system's ability to accomplish its computations at a sufficient speed to keep pace with the environment within which it is embedded. Rather, a real-time learning mechanism, as defined here, is one for which the time of occurrence of each critical event in the sequence leading to learning is of fundamental importance with respect to the computations the learning mechanism is performing. Real-time learning mechanisms may

be contrasted with nonreal-time learning mechanisms such as the perceptron (Rosenblatt, 1962), adaline (Widrow, 1962), or back propagation (Werbos, 1974; Parker, 1982, 1985; Le Cun, 1985; Rumelhart, Hinton, and Williams, 1985, 1986) learning mechanisms for which error signals follow system responses and only the order of the inputs, outputs, and error signals is important, not the exact time of occurrence of each signal, relative to the others. For additional discussions of real-time learning mechanism models, see Klopf (1972, 1975, 1979, 1982, 1986), Moore and Stickney (1980), Sutton and Barto (1981, 1987), Wagner (1981), Grossberg (1982, 1987), Schmajuk and Moore (1985), Gelperin, Hopfield, and Tank (1985), Blazis, Desmond, Moore, and Berthier (1986), Tesauro (1986), and Donegan and Wagner (1987). Proposals for real-time models that give especially careful attention to neurobiological constraints are those of Hawkins and Kandel (1984) and Gluck and Thompson (1987).

Klopf (1972, 1982) proposed an extension to Hebb's model that introduced the notions of synaptic eligibility and reinforcement into real-time learning mechanisms, resulting in a neuronal model that emphasized sequential rather than simultaneous events. The following sequence of critical events was hypothesized to lead to learning:

$$x_i(t-k) \rightarrow y(t-k) \rightarrow s(t) \rightarrow \Delta w_i(t) \qquad \text{(S-2)}$$

where $s(t)$ is the sum of the weighted inputs to the neuron at time t and k is the nominal interval of time required for a neuronal output to feed back and influence the neuronal input, the feedback occurring either through the remainder of the neural network or through the environment. The variable $s(t)$ represents the neuronal membrane potential. In this

model, presynaptic and postsynaptic activity, $x_i(t-k)$ and $y(t-k)$, when they occur in conjunction, render a synapse eligible for modification. However, the efficacy of an eligible synapse does not change unless the subsequent membrane potential, $s(t)$, is nonzero, $s(t)$ functioning as a reinforcer that follows the eligibility computation. The equation for the learning mechanism is as follows:

$$\Delta w_i(t)=cx_i(t-k)y_i(t-k)s(t) \qquad (4)$$

In the context of real-time learning mechanisms, the notions of synaptic eligibility and reinforcement based on sequential rather than simultaneous events yielded a neuronal model that could make greater contact with the experimental evidence of classical and instrumental conditioning (Klopf, 1972, 1982). A further step was taken in that direction when Barto and Sutton (1981a) discovered that replacing $s(t)$ in sequence (S-2) above with $\Delta s(t)$ permitted the neuronal model to make much more substantial contact with classical conditioning phenomena. The resulting neuronal learning mechanism is described by the following critical event sequence:

$$x_i(t-k) \rightarrow y(t-k) \rightarrow \Delta s(t) \rightarrow \Delta w_i(t) \qquad (S\text{-}3)$$

where $\Delta s(t) = s(t) - s(t-1)$.

The equation for the learning mechanism is:

$$\Delta w_i(t)=cx_i(t-k)y(t-k)\,\Delta s(t) \qquad (5)$$

This form of learning mechanism led to a simplification. Barto and Sutton (1981a) found that the critical event sequence (S-3) could be replaced with the following simpler sequence:

$$x_i(t-k) \rightarrow \Delta y(t) \rightarrow \Delta w_i(t) \qquad (S\text{-}4)$$

$\Delta y(t)$ in sequence (S-4) replaces $\Delta s(t)$ in sequence (S-3). This

can be seen to be plausible in that $\Delta y(t)$ implies $\Delta s(t)$. However, proceeding from sequence (S-3) to sequence (S-4) involved the additional discovery that $y(t-k)$ in sequence (S-3) was not essential for predicting classical conditioning phenomena. The result was a neuronal model that can be specified by the following equation:

$$\Delta w_i(t)=cx_i(t-k)\ \Delta y(t) \qquad (6)$$

Actually, the form the model took in the computer simulations Sutton and Barto (1981) reported was as follows:

$$\Delta w_i(t)=c\bar{x}_i(t)\ \Delta y(t) \qquad (7)$$

where

$$\bar{x}_i(t)=\ \alpha\bar{x}_i(t-1)+x_i(t-1) \qquad (8)$$

In equation (8), α is a positive constant. It can be seen that equation (7) is of a form similar to equation (6) except that $x_i(t-k)$ is replaced by $\bar{x}_i(t)$. $\bar{x}_i(t)$ represents an exponentially decaying trace of x_i extending over a number of time steps.

It was at this point that neuronal modeling intersected strongly with the theoretical and experimental results of animal learning researchers such as Kamin (1968) and Rescorla and Wagner (1972). Sutton and Barto (1981) demonstrated that the model they proposed could be seen as a temporally refined extension of the Rescorla-Wagner (1972) model. Like the Rescorla-Wagner model, the Sutton-Barto model accounted for a variety of classical conditioning phenomena including blocking, overshadowing, and conditioned inhibition. Here was what could be interpreted as a <u>neuronal</u> model (although Sutton and Barto did not insist on that interpretation) making predictions similar to those of a <u>whole animal</u> model! The Sutton-Barto model represented a milestone in terms of

the contact prospective neuronal models were making with the experimental evidence of animal learning (Sutton and Barto, 1981; Barto and Sutton, 1982; Moore, Desmond, Berthier, Blazis, Sutton, and Barto, 1986; Blazis and Moore, 1987).

However, the Sutton-Barto model still deviated from the experimental evidence in a number of significant respects. One problem was that the sensitivity of the model to conditioned stimulus durations caused the model to yield inaccurate predictions for a variety of conditioned stimulus-unconditioned stimulus configurations for which the conditioned stimulus and unconditioned stimulus overlapped significantly. The model also does not account for the initial positive acceleration in the s-shaped acquisition curves observed in classical conditioning.

One approach to correcting the problems of the Sutton-Barto model has been to utilize a variant of the adaptive heuristic critic algorithm developed by Sutton (1984), and this has led to the temporal difference model proposed by Sutton and Barto (1987). Temporal difference models, as defined by Sutton and Barto (1987), utilize differences between temporally successive predictions as a basis for learning. Sutton (1987) notes that the earliest and most well known use of a temporal difference (TD) method or model was that due to Samuel (1959) in his checker-playing program. Other examples of TD methods or models include those due to Witten (1977), Sutton and Barto (1981), Booker (1982), Hampson (1983/1984), Sutton (1984), Gelperin, Hopfield, and Tank (1985), and Holland (1986). The drive-reinforcement neuronal model proposed in this report is an example of a temporal difference model.

Variants of the adaptive heuristic critic model (Barto, Sutton, and Anderson, 1983; Sutton, 1984) represent one approach to solving the problems of the Sutton-Barto model. Seeking to address these same problems, I have adopted an alternative approach that has led to the neuronal learning mechanism specified by equation (2). For this model, the hypothesized sequence of critical events leading to learning is as follows:

$$\Delta x_i(t-j) \rightarrow \Delta y(t) \rightarrow \Delta w(t) \qquad \text{(S-5)}$$

where j replaces k and all of the critical events involve derivatives with respect to time. The variable, k, was the time required for the neuron to receive feedback regarding its earlier output, y(t-k); k reflected an instrumental conditioning orientation. The variable, j, is simply an interstimulus interval reflecting a classical conditioning orientation. Barto and Sutton had also considered using $\Delta x_i(t)$ instead of $x_i(t)$ in their learning mechanism but decided it was unworkable. I returned to this possibility of a <u>differential learning mechanism</u>, one that correlates earlier derivatives of inputs with later derivatives of outputs, and found a way to make it workable such that the problem with conditioned stimulus duration effects was eliminated. The class of differential learning mechanisms was independently discovered by Klopf (1986), coming from the directions of neuronal modeling and animal learning, and by Kosko (1986), coming from philosophical and mathematical directions.

Sequence (S-5) implies the following kind of learning mechanism:

$$\Delta w_i(t) = c\, \Delta x_i(t-j)\, \Delta y(t) \qquad (9)$$

However, I have found that the most workable form of the learning mechanism involves adding multiple terms and multiple learning rate constants to the right side of equation (9), the terms and constants corresponding to a range of interstimulus intervals, j. Also, making $\Delta w_i(t)$ proportional to the absolute value of $w_i(t-j)$ allows the model to account for the initial positive acceleration in the acquisition curves of classical conditioning. These refinements led to the neuronal learning mechanism specified by equation (2) and repeated here:

$$\Delta w_i(t) = \Delta y(t) \sum_{j=1}^{\tau} c_j \, |w_i(t-j)| \, \Delta x_i(t-j) \qquad (10)$$

where $\Delta x_i(t-j)$ must be greater than or equal to zero; otherwise, $\Delta x_i(t-j)$ is set equal to zero for the purposes of equation (10). The resulting model predicts the basic categories of classical conditioning phenomena, as will be demonstrated in the next section.

SECTION 3

CLASSICAL CONDITIONING: PREDICTIONS OF THE NEURONAL MODEL

Classical conditioning phenomena are basic to learning. I will show in this section that the drive-reinforcement neuronal model predicts a wide range of classical conditioning phenomena. This will be demonstrated by means of computer simulations of the model.

The neuronal model that was simulated is shown in Figure 3. The input-output (I/O) relationship assumed for the neuron was that of equation (1). The neuronal learning mechanism (L.M.) was that of equation (2) with the refinement noted earlier: whenever $\Delta x_i(t-j)$ was less than zero, $\Delta x_i(t-j)$ was set equal to zero for the purpose of calculating $\Delta w_i(t)$. In the computer simulations, a conditioned stimulus (CS) or unconditioned stimulus (US) that was presented to the neuron had an amplitude that ranged between zero and one and a duration that was specified in terms of the times of stimulus onset and offset. In the figures showing results of the computer simulations, each CS-US configuration is graphed so the reader may see the relative amplitudes and durations of stimuli at a glance. (For exact values for any of the parameters for the computer simulations, the Appendix should be consulted.)

Each stimulus was presented to the simulated neuron through both an excitatory and an inhibitory synapse so that the neuronal learning mechanism had, for each input, both an excitatory and an inhibitory weight available for modification. The learning mechanism could then

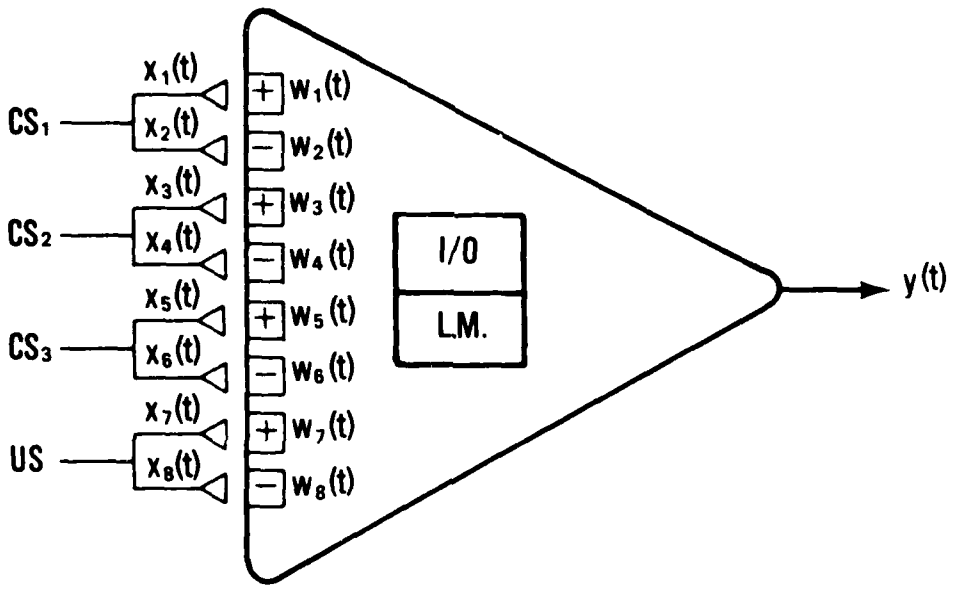


Figure 3. The drive-reinforcement neuronal model employed in the computer simulations. This is a specific example of the more general model shown in Figure 1. The description that was given in Figure 1 applies here. In addition, each CS and US is represented by an excitatory (+) and an inhibitory (-) synapse. The efficacies of synapses [i.e., the synaptic weights, $w_i(t)$] are variable (plastic) for synapses mediating CSs and fixed (nonplastic) for synapses mediating USs.

choose to modify one or the other weight or both in each time step. In the case of an actual (biological) neuron, if a CS is not represented by both excitatory and inhibitory synapses, the individual neuron will be constrained in terms of what classical conditioning phenomena it can manifest. It will be seen in the simulations below that, for a drive-reinforcement neuron, some classical conditioning phenomena require only excitatory plastic synapses and some require only inhibitory plastic synapses. Those classical conditioning phenomena requiring both excitatory and inhibitory plastic synapses would have to emerge at a higher level if the individual neurons involved had their CSs represented by only excitatory or only inhibitory plastic synapses.

In the discussion that follows, a conditioned or unconditioned stimulus and the associated $x_i(t)$ in Figure 3 are identical. For example, $x_1(t)$ and $x_2(t)$ are one and the same as CS_1. The weights associated with the synapses carrying the unconditioned stimulus were fixed (nonplastic) and the remaining synaptic weights were variable (plastic).

The conditioned stimulus or unconditioned stimulus that is described should, perhaps, more properly be referred to as a neuronal conditioned stimulus or a neuronal unconditioned stimulus because it is the stimulus that is reaching the neuron, not the stimulus that is reaching the whole animal. However, for the sake of simplicity in the discussion, I will refer to these neuronal input signals as conditioned and unconditioned stimuli or, simply, CSs and USs. Likewise, the output, $y(t)$, of the neuron would more properly be referred to as the neuronal conditioned or unconditioned response but I will usually refer to the neuronal response

as the conditioned response (CR) or unconditioned response (UR). Built into these terminological conventions is the assumption that stimuli and responses external to an animal's nervous system do not differ fundamentally in form from the way stimuli and responses are represented internal to the animal's nervous system. This assumption might not hold up well at higher, cognitive levels of function but the assumption appears reasonable as a starting point for testing the ability of a neuronal model to predict fundamental learning phenomena.

Just as the range of $x_i(t)$ in the simulations was from zero to one, as was noted when the range of CS and US amplitudes was discussed, so the range of $y(t)$, the neuronal output was from zero to one. Such a range serves to model a finite range of frequencies for neuronal inputs and outputs. Actual frequencies of biological neurons range up to several hundred spikes per second in the case of neocortical neurons firing for brief intervals (Lynch, Mountcastle, Talbot, and Yin, 1977). Therefore, one could multiply the neuronal input and output amplitudes used in the simulations by, say, three hundred if one desires to see more realistic numbers. However, for the purposes of the simulations to be reported, the relative magnitudes of the parameters are important, not the absolute magnitudes.

The number of synapses impinging on the simulated neuron is eight, as is indicated in Figure 3. This corresponds to three possible CSs and one US. The absolute values of the plastic synaptic weights mediating the CSs have a lower bound of 0.1 and, when the simulations began, these excitatory and inhibitory weights were set at plus and minus 0.1, respectively. (For exceptions to this statement, see the Appendix; in

some simulations, inhibitory synaptic weights were set equal to zero because they did not play a significant role and it simplified the graphs.) The neuronal threshold was set at zero because, at higher values of the neuronal threshold, the form of the model's predictions did not change. The only effect of higher thresholds was that more trials were required for the synaptic weights to reach their asymptotic values. For the learning mechanism, the learning rate constants, c_1 through c_5, were set at values such that $c_j > c_{j+1}$. As noted earlier, this is reasonable if one views each time step as being equivalent to one-half second because then c_1 is maximal, corresponding to a nominal optimal interstimulus interval of one-half second. Successive c-values then decrease as the interstimulus interval increases. As also noted earlier, c_0 and c_6 were set equal to zero, corresponding to interstimulus intervals of zero and three seconds, respectively. Thus, in the simulations, j ranged from one to five; i.e., τ was set equal to five.

What follows are the results of computer simulations of the drive-reinforcement neuronal model for a variety of CS-US configurations. The predictions of the model are examined for delay and trace conditioning, CS and US duration and amplitude effects, partial reinforcement effects, interstimulus interval effects including simultaneous conditioning, second-order conditioning, conditioned inhibition, extinction, reacquisition effects, backward conditioning, blocking, overshadowing, compound conditioning, and discriminative stimulus effects.

During a simulation, the CS-US configuration was presented once in each trial. The values of the synaptic weights at the end of each trial

were recorded and plotted as a function of the trial number. These graphs of synaptic weights versus trials are shown in the figures accompanying the discussion below. In addition, in each figure, the CS-US configuration is graphed along with the response of the neuron during the last trial. The neuronal response is labeled "Y," designating a plot of $y(t)$ for the last trial of the simulation. The definition of a trial should be noted. The CS-US configuration, or what is referred to in the figures as the "stimulus configuration", defines a trial. Thus, the graphed stimulus configurations in the figures are intended to show not only relative times of onset and offset along with amplitudes of stimuli but also the number of times a stimulus was presented during a trial. What will be seen in the figures is that the behavior of the synaptic weights, as predicted by the drive-reinforcement neuronal model, mirrors the observed behavior of animals as they are learning during classical conditioning experiments.

Before discussing the individual simulations, two remarks are in order regarding the graphs of synaptic weights versus trials. Any synaptic weight that played a significant role for the conditioning phenomenon being discussed is shown in the accompanying graph. Any synaptic weight that played no significant role (typically meaning that the neuronal learning mechanism did not alter the weight at all during the simulation) is not shown in order to simplify the graphs. Also, data points for the synaptic weight values at the end of each trial are not shown on the graphs because the resulting density of the data points would be excessive and because the data points fall exactly on the (theoretical) curves that have been drawn.

Delay Conditioning

Delay conditioning is defined such that CS onset precedes US onset and CS offset occurs at the same time as or after US onset. An example is the well known Pavlovian experiment in which a bell (the CS) is paired with food (the US). The observed result in such experiments is that conditioned excitation develops. The bell becomes excitatory with respect to the salivary gland. In addition, it is observed that the amount of salivation in response to the bell alone (measured with occasional test probes) increases with increasing trials such that an s-shaped or sigmoid curve results when the amount of salivation is plotted versus the trial number. That is to say, the amount of salivation in response to the bell alone, as a function of trials, positively accelerates initially and then negatively accelerates as an asymptotic level of conditioning is approached (Pavlov, 1927). Spence (1956) has observed that the acquisition curves of classical conditioning are always s-shaped, providing that the experiments are done carefully enough to capture the initial positive acceleration and the later negative acceleration. For example, Spence (1956, pp. 68-70) states that acquisition curves that "do not exhibit an initial, positively accelerated phase do not do so either because they do not start at zero level of conditioning or because the conditioning is so rapid that the period of initial acceleration is too brief to be revealed except by very small groups or blocks of trials."

Figure 4 shows the predicted acquisition curves of three neuronal models for delay conditioning. In Figure 4(a), the results of a simulation of the model proposed by Hebb (1949) are shown. For the

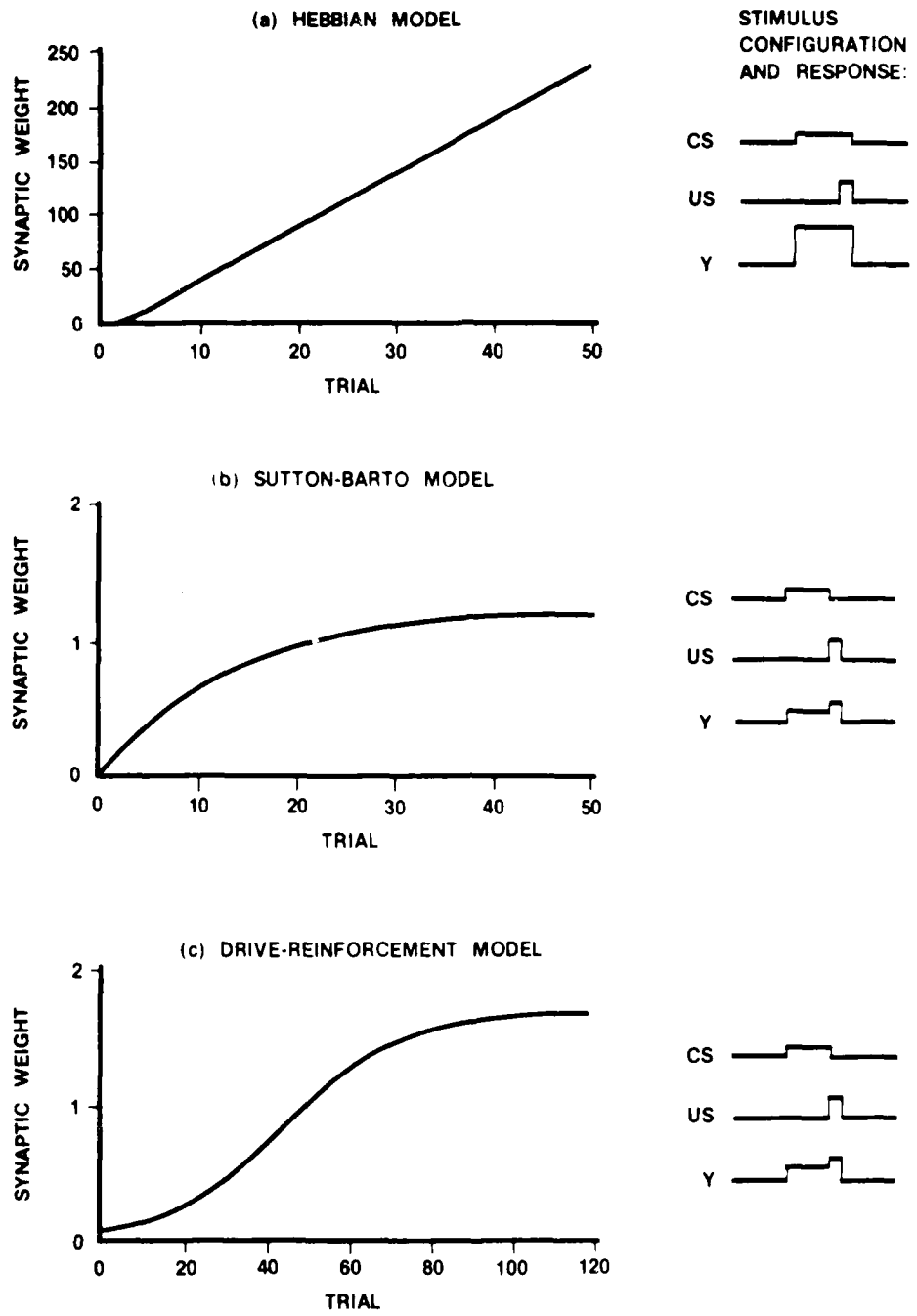


Figure 4. Results of simulated delay conditioning experiments with (a) Hebbian, (b) Sutton-Barto, and (c) drive-reinforcement neuronal models. The Hebbian model yields an essentially linear acquisition curve. The Sutton-Barto model yields a negatively accelerated acquisition curve. Consistent with the experimental evidence, the drive-reinforcement neuronal model yields an s-shaped acquisition curve. (See text and Appendix for details.)

Hebbian model, the input-output relationship is the same as for the drive-reinforcement model and is, therefore, specified by equation (1). The Hebbian learning mechanism has already been noted and is specified by equation (3). It can be seen in Figure 4(a) that if a Hebbian neuron were driving the salivary gland, the amount of saliva produced in response to the bell alone as a function of trials would exhibit an essentially linear relationship because the excitatory synaptic weight associated with the CS varies in an essentially linear fashion with the trial number. Also, it may be noted that the Hebbian learning mechanism does not yield an asymptotic synaptic weight value but, rather, continues to increase the synaptic weight indefinitely or, of course, until an upper bound would be reached.

In Figure 4(b), the results of a simulation of the Sutton-Barto (1981) model are shown. The Sutton-Barto learning mechanism was specified earlier in equations (7) and (8). The model's input-output relationship is that of equation (1). The model is seen to predict a negatively accelerated acquisition curve in that the excitatory synaptic weight associated with the CS negatively accelerates with increasing trials. It may be noted that the Rescorla-Wagner (1972) model also predicts a negatively accelerated acquisition curve, as have earlier whole animal models [see, for example, a model due to Estes (1950)].

In Figure 4(c), the results of a simulation of the drive-reinforcement model are shown. The model is seen to predict an s-shaped acquisition curve: Conditioned excitation develops, first through a positively accelerating phase and then through a negatively accelerating phase. The drive-reinforcement model is

thus seen to be consistent with this aspect of the experimental evidence of delay conditioning.

Some reasons why the drive-reinforcement model yields an s-shaped acquisition curve may be noted. The initial positive acceleration is due to the efficacy of the relevant synapse appearing as a factor on the right side of equation (2). Thus, as the learning mechanism increases the efficacy of the synapse, the future rate of change of the efficacy of the synapse is also caused to increase. With continued conditioning, another process comes to dominate, yielding the eventual negative acceleration in the acquisition curve. The negative acceleration is due to $\Delta y(t)$ decreasing with continued conditioning. In effect, $\Delta y(t)$ moves to an earlier point in time with conditioning, becoming $\Delta y(t-j)$ where j is the interstimulus interval. Thus, throughout the conditioning process, increasing values of $w_i(t-j)$ are competing with decreasing values of $\Delta y(t)$ in equation (2). Rapidly increasing values of $w_i(t-j)$ prevail initially and rapidly decreasing values of $\Delta y(t)$ prevail later, yielding the respective positive and negative accelerations in the acquisition curve.

CS and US duration effects

A careful reader may note that, in Figure 4, the same CS-US configuration is not used for the simulation of each of the models. The Hebbian model's CS offset coincides with the offset of the US whereas the Sutton-Barto and drive-reinforcement model's CSs have the offset occuring at the time of US onset. I chose those particular CS-US configurations because, otherwise, the Hebbian and Sutton-Barto models would not have

predicted the development of conditioned excitation. Both of these models are sensitive to CS durations in a way that is not consistent with the experimental evidence, the models predicting no conditioning or conditioned inhibition for some CS-US configurations that, experimentally, are known to yield conditioned excitation. The effect of CS duration is examined systematically in Figure 5 where each model's predictions are shown for the same set of three CS-US configurations. I will specify how the three CS-US configurations differ and then discuss each model's predictions for each of the three configurations.

In Figure 5, CS_1 offset occurs at the time of US onset, CS_2 offset occurs at the time of US offset, and CS_3 offset occurs one time step after US offset. Experimentally, it is known that conditioned excitation (corresponding in the neuronal models to the growth of positive synaptic weights) is observed in all three cases. In general, the efficacy of delay conditioning is a strong function of the time of CS onset and relatively independent of CS duration (Kamin, 1965).

In Figure 5(a), it is seen that the Hebbian model predicts conditioned excitation for CS_2 and CS_3 but not for CS_1. In Figure 5(b), it is seen that the Sutton-Barto model predicts conditioned excitation for CS_1 and strong conditioned inhibition for CS_2 and CS_3. In Figure 5(c), it is seen that, consistent with the experimental evidence, the drive-reinforcement model predicts conditioned excitation for all three CSs and, in each case, predicts an s-shaped acquisition curve.

In Figure 5(c), more detailed aspects of the drive-reinforcement model's predictions may be noted. For example, the model predicts a

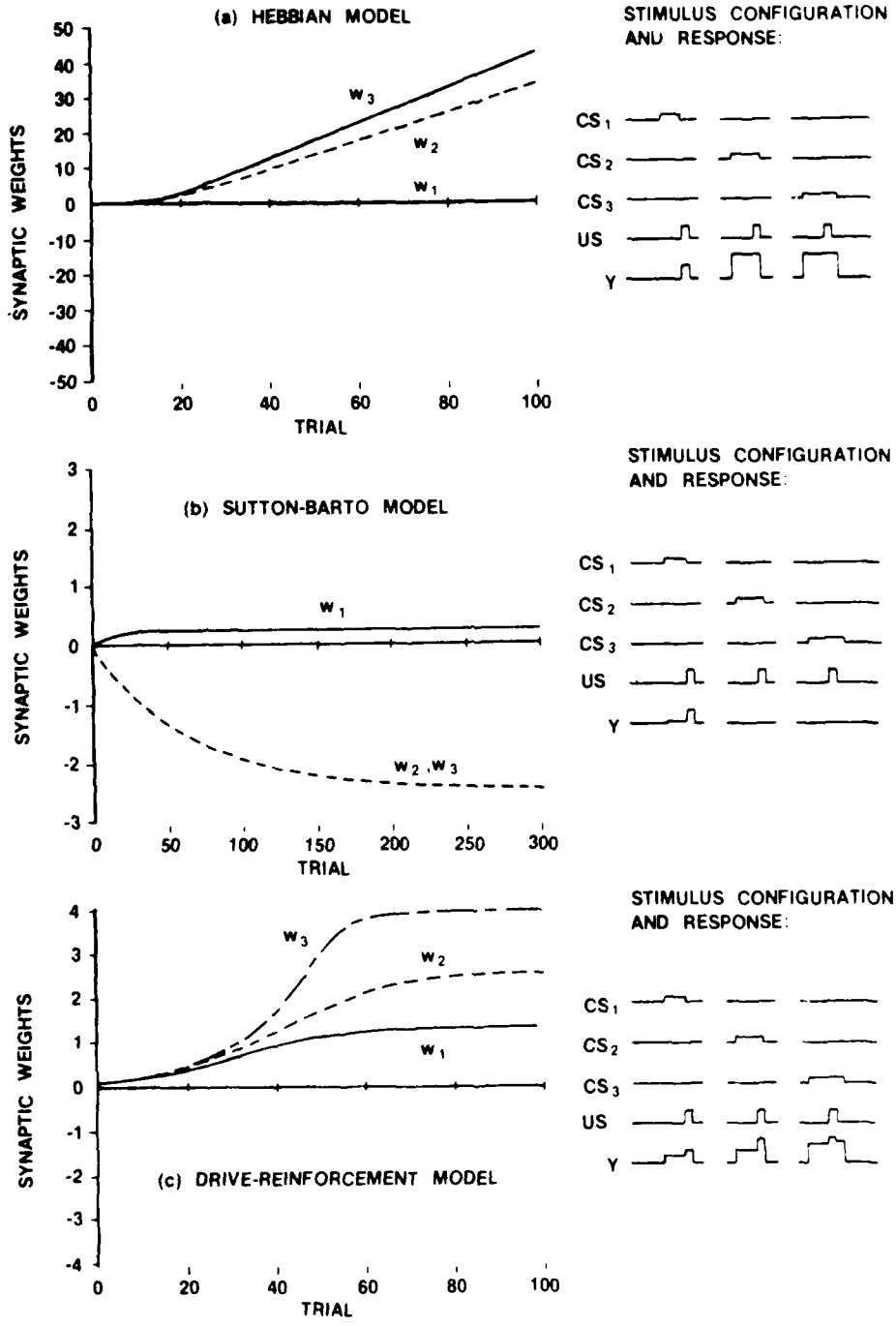
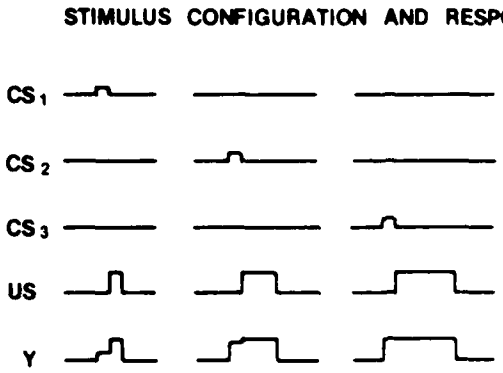


Figure 5. Results of simulated delay conditioning experiments with (a) Hebbian, (b) Sutton-Barto, and (c) drive-reinforcement neuronal models. The effect of CS duration is examined. (See text and Appendix for details.)

particular ranking of CSs in terms of initial rate of conditioning and asymptotic synaptic weight value as a function of CS duration. The experimental literature does not, at this point, permit the accuracy of these more detailed predictions to be assessed. Furthermore, whole animal data may be insufficient to test these predictions, in that higher level attention mechanisms may play a significant role when CS durations are extended beyond the US (Ayres, Albert, and Bombace, 1987). Experiments at the level of the single neuron may be required to test these predictions.

Regarding the effects of US duration, the drive-reinforcement model predicts increasing rates of conditioning as the US duration increases (see Figure 6) and this is consistent with the experimental evidence (Ashton, Bitgood, and Moore, 1969; Gormezano, Kehoe, and Marshall, 1983).

Thus far, the drive-reinforcement neuronal model's predictions have been demonstrated to be accurate for three categories of classical conditioning phenomena: (a) the form of the acquisition curve in delay conditioning, (b) relative insensitivity to CS duration, and (c) US duration effects. The predictions of the model for a variety of other CS-US configurations will now be examined, these CS-US configurations corresponding to what appear to be the remaining basic categories of classical conditioning phenomena. While the predictions of the Hebbian and Sutton-Barto models for these CS-US configurations will not be shown, it should be noted that the Hebbian model's predictions frequently deviate substantially from experimentally observed behavior, examples of this having already been seen in Figures 4 and 5. (Of course, it remains a theoretical possibility that biological neurons are Hebbian and that



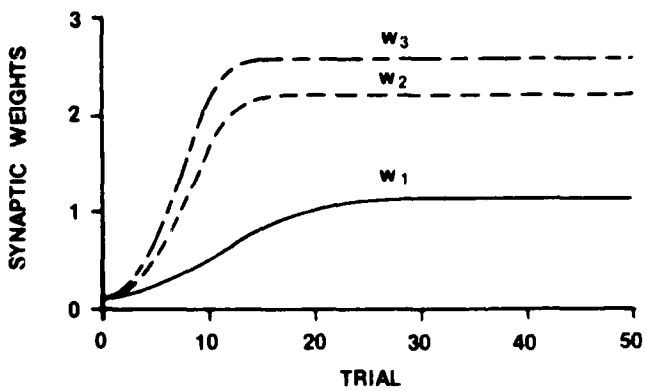
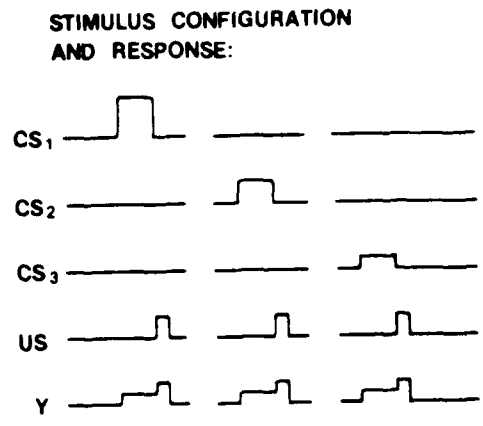


Figure 6. The drive-reinforcement model's predictions of the effects of US duration. Consistent with the experimental evidence, as the US duration increases, the excitatory synaptic weights associated with the reinforced CSs increase more rapidly and reach a higher asymptotic level. (See text and Appendix for details.)

classical conditioning phenomena are emergent, resulting from the interactions of perhaps large numbers of Hebbian neurons. Experimental tests to be discussed later will be required to resolve this question.) The predictions of the Sutton-Barto model are similar to those of the drive-reinforcement model, if one is careful, in the case of the Sutton-Barto model, not to use substantially overlapping CSs and USs and accepting that the Sutton-Barto model's predicted acquisition curves are not s-shaped.

CS and US amplitude effects

It is known that faster conditioning results as the intensity of the CS increases (Pavlov, 1927; see review by Moore and Gormezano, 1977). As is seen in Figure 7, the drive-reinforcement model predicts this relationship. Shown in Figure 7 are CSs of three different amplitudes, each being reinforced by a US of the same amplitude. The predicted rate of conditioning is seen to increase as the amplitude or intensity of the CS increases. For the three CSs, the rank ordering of the asymptotic values of the synaptic weights is the reverse of the rank ordering of the rates of acquisition because a low amplitude CS requires a larger asymptotic synaptic weight to yield the same eventual CR amplitude as can be obtained with a high amplitude CS and a lower asymptotic synaptic weight.



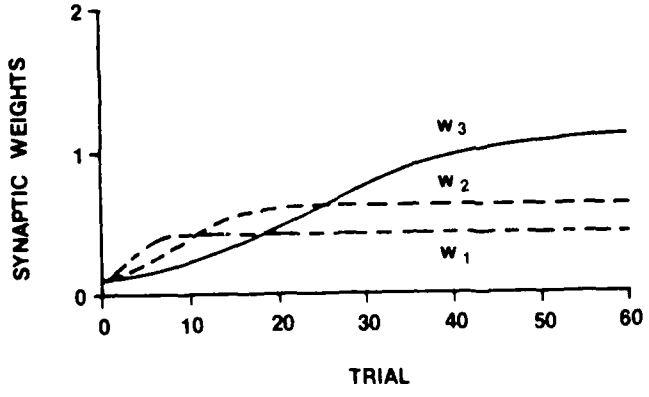


Figure 7. The drive-reinforcement model's predictions of the effects of CS amplitude. Consistent with the experimental evidence, as the CS amplitude decreases, the rate of growth of the excitatory synaptic weights associated with the reinforced CSs decreases. Asymptotic excitatory synaptic weight values vary inversely with CS amplitude because a lower CS amplitude requires a higher excitatory asymptotic synaptic weight value to yield a CR amplitude equal to the UR amplitude. (See text and Appendix for details.)

Regarding US amplitude effects, Moore and Gormezano (1977, p. 115) note that "Within limits, the rate of acquisition and level of performance of a CR are increasing functions of the intensity of the US." This is predicted by the drive-reinforcement model as can be seen in Figure 8 where three identical CSs are shown being reinforced by USs of decreasing amplitude. It is seen that both the rate of acquisition and the asymptotic weight value decrease as the US amplitude decreases.

CS preexposure effects

CS preexposure refers to nonreinforced presentations of a CS prior to reinforced presentations. The observed result is that CS preexposure retards subsequent acquisition of the conditioned response when reinforced presentations of the CS begin but the experimental evidence also suggests that the preexposed CS does not become inhibitory [see review by Flaherty (1985) who cites, e.g., Rescorla (1971), Reiss and Wagner (1972) and Solomon, Brennan and Moore (1974)]. As Flaherty (1985) notes, one possible explanation for CS preexposure effects is that the animal may, during the nonreinforced CS presentations, learn not to attend to the stimulus. If this is the case, CS preexposure effects would not be predicted by a neuronal model. Rather, such effects would require network-level considerations for their prediction. The related subject of US preexposure effects will be discussed later when the phenomenon of blocking is considered.

Partial reinforcement effects

In the case of partial reinforcement, a CS is not always followed by a US. This can be contrasted with continuous reinforcement, in which case the US always follows the CS. The observed result of partial

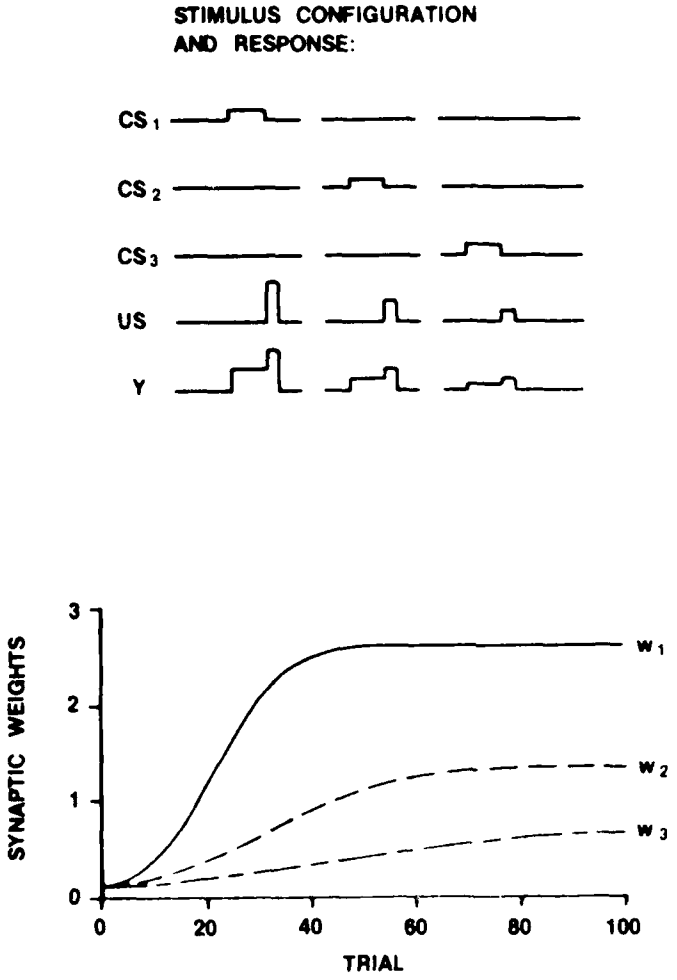


Figure 8. The drive-reinforcement model's predictions of the effects of US amplitude. Consistent with the experimental evidence, as the US amplitude decreases, the rates of growth and asymptotic values of excitatory synaptic weights associated with the reinforced CSs decrease. (See text and Appendix for details.)

reinforcement is a reduced rate of conditioning and sometimes a reduced asymptotic level of responding (Gormezano, Kehoe, and Marshall, 1983) relative to the rates and asymptotic levels observed for continuous reinforcement. The drive-reinforcement model's predictions are consistent with this, as can be seen in Figure 9, where CS_1 is reinforced 100 percent of the time, CS_2 is reinforced 50 percent of the time, and CS_3 is reinforced 25 percent of the time. In Figure 9, it is seen that rates of acquisition and asymptotic weight values are predicted to decrease as the percent reinforcement decreases.

Trace conditioning

Trace conditioning is an experimental procedure in which CS offset precedes US onset. The time between CS offset and US onset is termed the trace interval. In general, the longer the trace interval, the lower the rate of acquisition and the lower the asymptotic level of conditioning [See Flaherty (1985) for a review of the experimental evidence]. The drive-reinforcement model predicts these relationships, as can be seen in Figure 10, where three CS-US configurations are shown. It can be seen that increasing trace intervals yielded both lower rates of acquisition and lower asymptotic synaptic weight levels. In terms of the drive-reinforcement model's dynamics, some reasons that trace conditioning is less effective than delay conditioning are that the Δx that occurs at CS onset is paired not only with the positive Δy of US onset but also with the negative Δy of CS offset and, furthermore, the interstimulus interval for the negative Δy has a larger learning rate constant associated with it than does the interstimulus interval for the positive Δy.

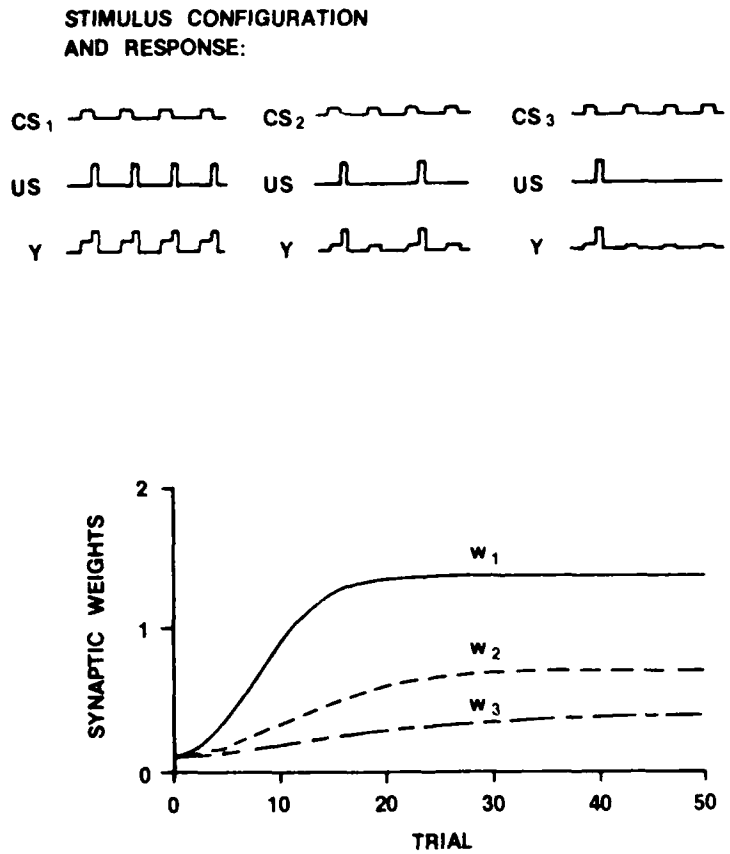


Figure 9. The drive-reinforcement model's predictions of the effects of partial reinforcement. Consistent with the experimental evidence, it is seen that as the fraction of CSs that are reinforced decreases, so does the rate of growth of excitatory synaptic weights associated with the reinforced CSs. The drive-reinforcement model also predicts lower asymptotic excitatory synaptic weight values as the percentage of reinforced CSs decreases, an effect that is consistent with some partial reinforcement studies. (See text and Appendix for details.)

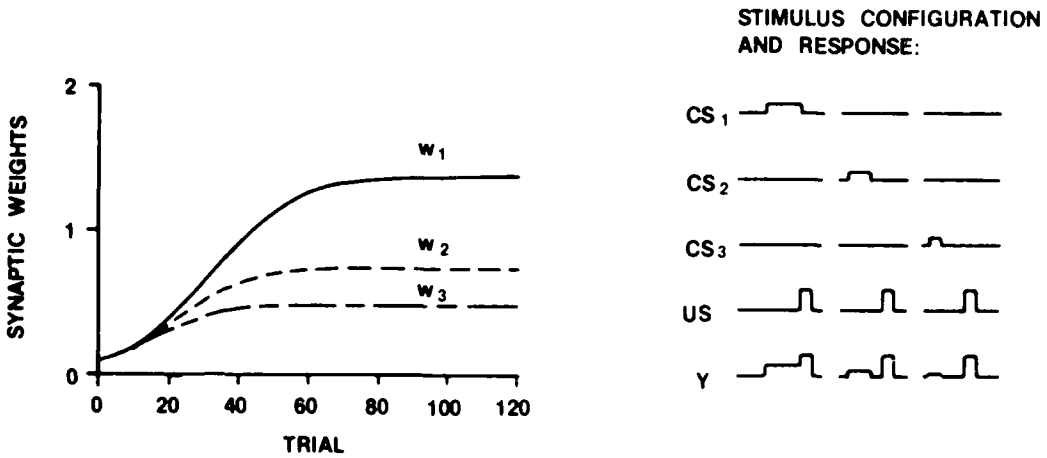


Figure 10. The drive-reinforcement model's predictions of the effects of trace conditioning. Consistent with the experimental evidence, as the trace interval increases, the rates of growth and asymptotic values of the excitatory synaptic weights associated with the reinforced CSs decrease. (See text and Appendix for details.)

Interstimulus interval effects including simultaneous conditioning

The predictions of the drive-reinforcement model for a variety of interstimulus intervals in delay conditioning are shown in Figure 11. The interstimulus interval is defined to be the time between CS and US onsets. In the case of CS_1 in Figure 11, CS and US onsets are simultaneous. This CS-US configuration is an example of what is referred to as simultaneous conditioning. Citing Pavlov (1927) as well as Smith, Coleman, and Gormezano (1969), Flaherty (1985) notes that "little or no conditioning occurs with simultaneous CS and US onset." This is what the drive-reinforcement model predicts. As can be seen in Figure 11, the synaptic weight for CS_1 remains unchanged during the sixty trials for which the computer simulation was run. Flaherty (1985) goes on to note that some conditioning has been reported for simultaneous CS and US onsets in the case of fear conditioning (Burkhardt and Ayres, 1978; Mahoney and Ayres, 1976). Thus, the experimental results with regard to simultaneous conditioning appear complex and it can only be noted that the predictions of the drive-reinforcement model appear to be consistent with some of the experimental evidence.

For interstimulus intervals greater than zero, experimental results suggest that a nominal interval of 500 ms (one time step in the simulations) is optimal when conditioning short latency skeletal reactions. With longer intervals, the efficacy of conditioning declines until, for intervals exceeding a few seconds, no conditioning is observed (see review by Moore and Gormezano, 1977). This is consistent with the predictions of the drive-reinforcement model. In Figure 11, it is seen that conditioning is most rapid for an interstimulus interval of one time

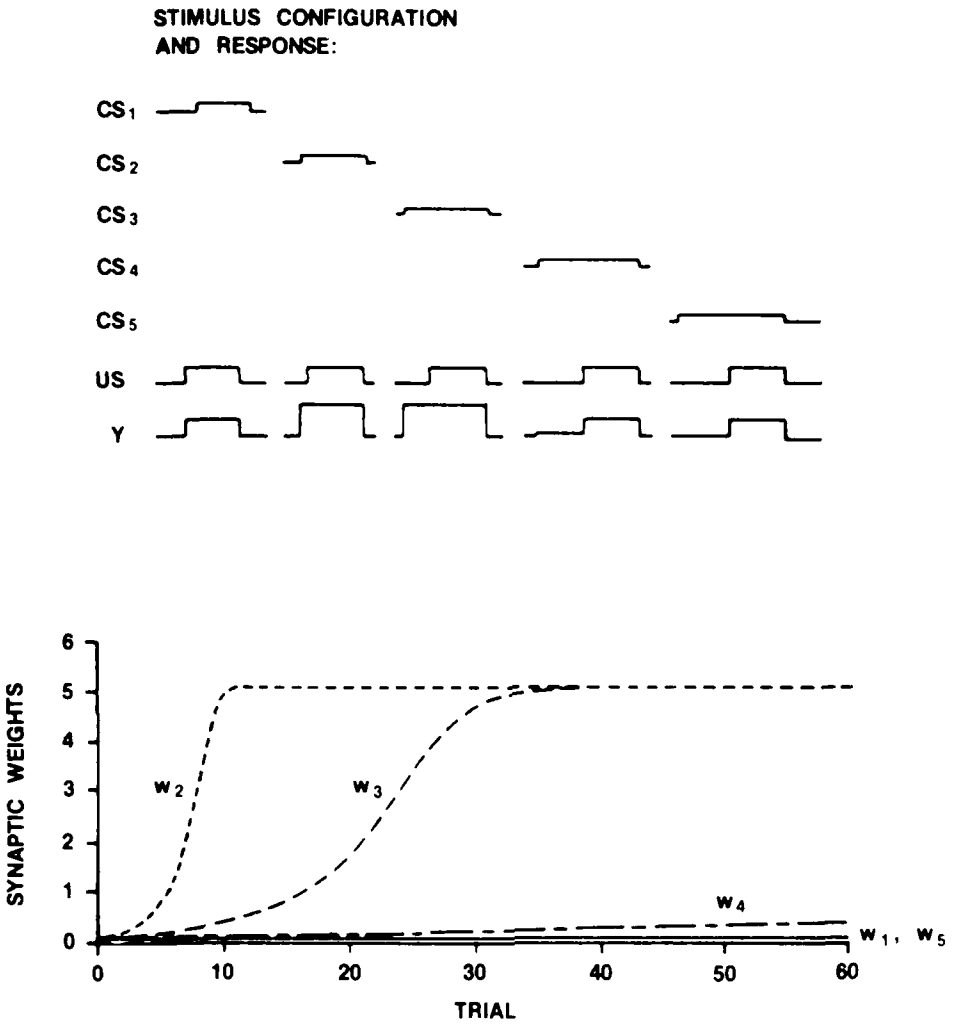


Figure 11. The drive-reinforcement model's predictions of the effect of the interstimulus interval. Consistent with the experimental evidence and consistent with the assignment of values to the learning rate constants, c_i, the model is seen to predict no conditioning for simultaneous CS_1 and US onsets and then decreased rates of conditioning as the interstimulus interval increases beyond the optimal interstimulus interval employed with CS_2. Interstimulus intervals were as follows: zero time steps for CS_1, one time step for CS_2, three time steps for CS_3, five time steps for CS_4, and six time steps for CS_6. (See text and Appendix for details.)

step in the case of CS_2, progressively slower for intervals of three and five time steps in the cases of CS_3 and CS_4, respectively, with no conditioning manifesting for an interstimulus interval of six time steps in the case of CS_5.

An alternative way of viewing the simulation results shown in Figure 11 is to see them as confirming the expected consequences of assigning the learning rate constants, c_j, in the manner described earlier. Namely, c_0 and c_6 were set equal to zero, c_1 was assigned the highest value and c_2 through c_5 were assigned progressively lower values. Thus, the simulation results in Figure 11 reflect the fact that the learning rate constants were chosen consistent with the empirical evidence regarding interstimulus interval effects.

Second-order conditioning

Second-order conditioning is an experimental procedure in which one CS is reinforced by another CS, the latter CS having been previously reinforced by a US. Pavlov (1927) reported that this procedure yielded conditioning in the second stage, the second CS coming to elicit the conditioned response originally elicited only by the first CS. However, in discussing second-order conditioning, Rescorla (1980, pp. 3-4) comments on "a historically nagging issue". Rescorla states that the "issue concerns whether, in fact, second-order conditioning is a real and powerful phenomenon. Although Pavlov reported its occurrence, he described it as transient. Subsequent authors have often been less than enthusiastic about its reality." This is interesting because the drive-reinforcement model predicts that second-order

conditioning will not be as strong as first-order conditioning and that second-order conditioning will be transient. Simulation results that are the basis of this prediction are shown in Figure 12 where, in stage one of conditioning, CS_1 is reinforced by a US, achieving an asymptotic synaptic weight value of just a little more than four. After delay conditioning in stage one (trials 1-60), second-order conditioning occurs in stage two (trials 61-200). The drive-reinforcement model predicts significantly weaker conditioning in stage two, the synaptic weight associated with CS_2 peaking at a value between one and two. Furthermore, the transient nature of second-order conditioning, as reported by Pavlov (1927), is predicted by the model. In stage two of the simulated second-order conditioning experiment, after the CS_2 synaptic weight peaks, the model predicts the subsequent decline of the weight due to what is essentially an extinction process. Had the simulation been carried out for further trials, the CS_2 synaptic weight would have declined to the lower bound of 0.1.

Conditioned inhibition

Delay conditioning yields conditioned excitation; i.e., the CS comes to excite the conditioned response (CR). An alternative procedure developed by Pavlov (1927) yields what he termed conditioned inhibition; i.e., a CS would come to inhibit a CR that otherwise would have manifested.

One of Pavlov's procedures for demonstrating conditioned inhibition was as follows. In the first stage of conditioning, Pavlov would utilize a delay conditioning procedure to render CS_1 excitatory with respect to a CR. Then, in a second stage of

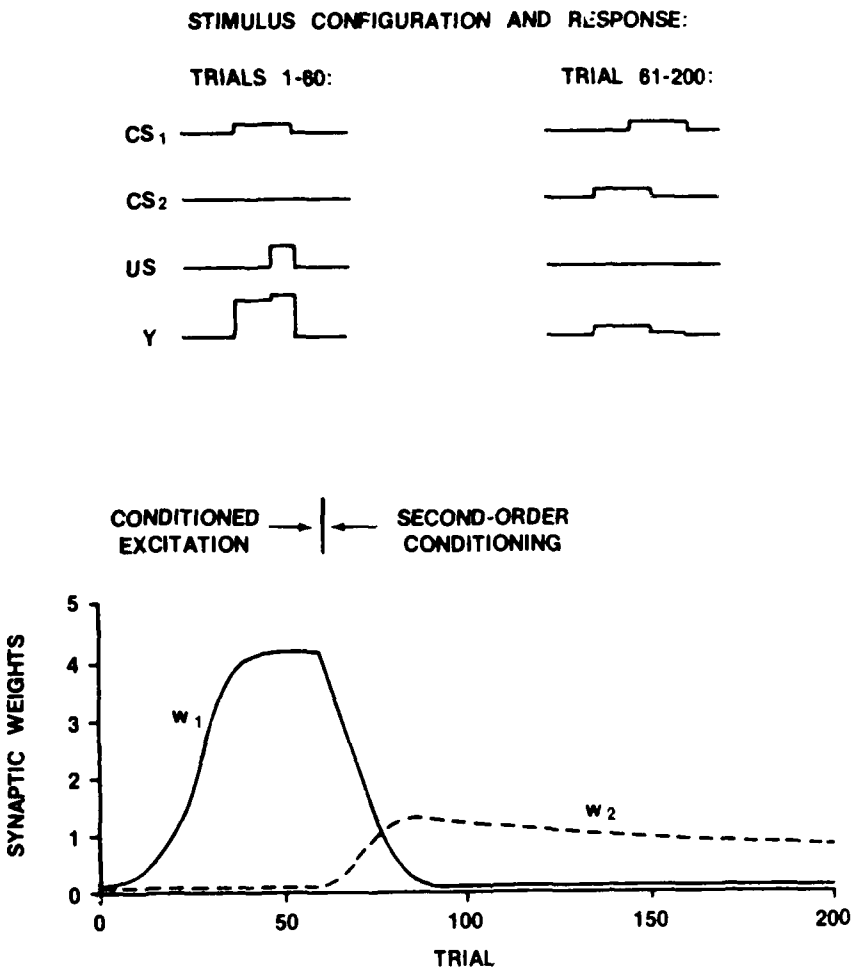


Figure 12. The drive-reinforcement model's predictions of the effects of second-order conditioning. Consistent with the experimental evidence, after delay conditioning in stage 1 (trials 1-60), the excitatory synaptic weight associated with CS_1 extinguishes in stage 2 (trials 61-200) during second-order conditioning. Also consistent with the experimental evidence, the excitatory synaptic weight associated with CS_2 increases initially during stage 2 and then decreases. (See text and Appendix for details.)

conditioning, he would continue to reinforce CS_1 with the US but he would also present an unreinforced CS_1-CS_2 pair to the animal. During the second stage of conditioning, the animal's response to CS_1 unpaired would decrease initially and then return to its original level. The animal's response to the CS_1-CS_2 pair would decrease to zero. Furthermore, Pavlov was able to demonstrate that CS_2 became a conditioned inhibitor in that, after stage-two conditioning, if CS_2 was paired with another CS, say CS_3, that was known, by itself, to be a conditioned exciter, the CR associated with CS_3 was, in general, reduced or eliminated.

The drive-reinforcement model predicts this behavior, as can be seen in Figure 13. In stage one (trials 1-70) of the simulated conditioning, CS_1 is reinforced by a US such that conditioned excitation develops, with the progress of the excitatory weight, $w_1(E)$, exhibiting the usual s-shaped acquisition curve. Then, in stage two (trials 71-200), CS_1 unpaired is reinforced by the US once in each trial while the CS_1-CS_2 pair is also presented once during each trial and the pair is unreinforced. The model predicts that the excitatory weight associated with CS_1 will decrease initially and then return to its previous level, mirroring the behavior Pavlov observed with his animals. Also, the model predicts that the inhibitory weight, $w_2(I)$, associated with CS_2, will grow stronger as stage two conditioning proceeds, consistent with Pavlov's observation that CS_2 becomes a conditioned inhibitor. (Regarding the notation employed here, an "E" or an "I" in parentheses following "w_i" signifies an excitatory or inhibitory weight, respectively. This notation involves a degree of redundancy in that excitatory weights will always be positive and inhibitory weights will

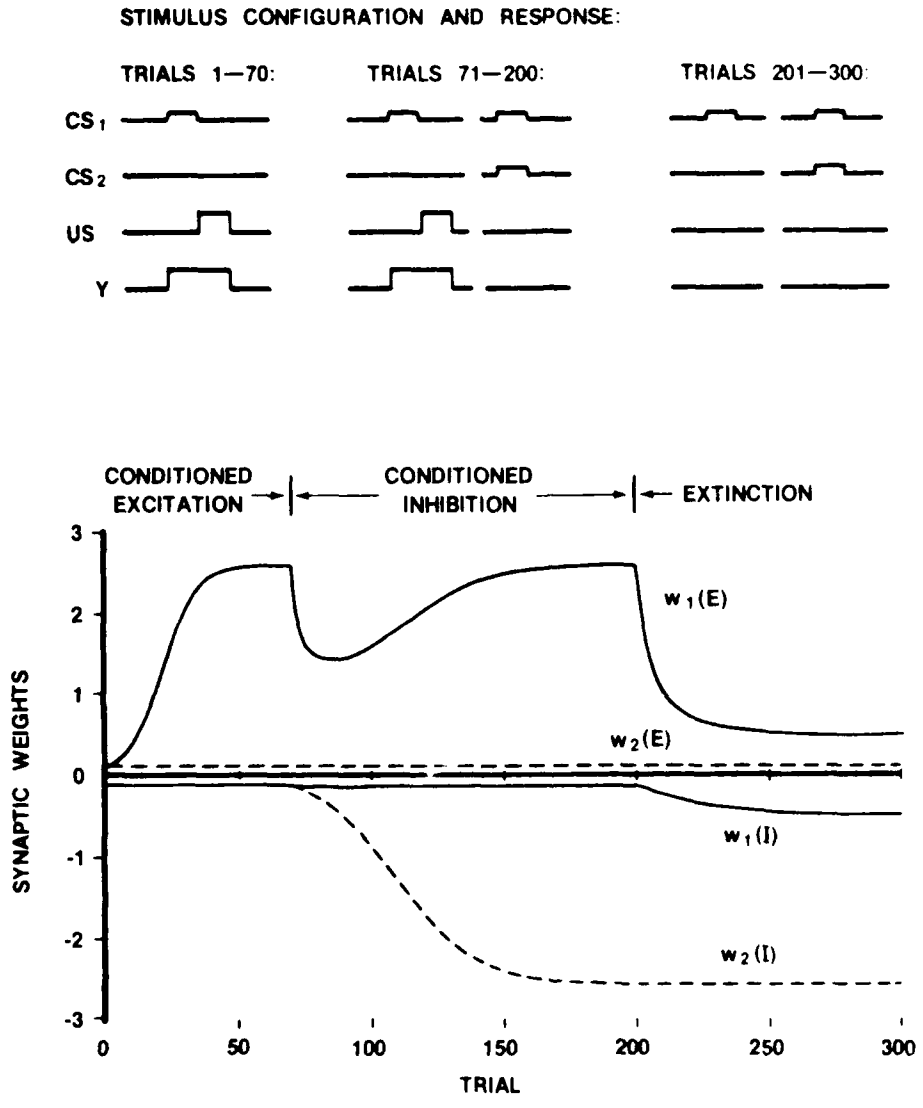


Figure 13. Results of a simulated classical conditioning experiment modeled after experiments performed by Pavlov (1927), in which conditioned excitation, conditioned inhibition, and extinction paradigms are employed. (See text and Appendix for details.)

always be negative so, in the graphs, excitatory and inhibitory weights for a particular CS could be distinguished on that basis.)

Because the decrease in the excitatory weight associated with CS_1 during the second stage of conditioning and then its subsequent return to the asymptotic level achieved in the first stage of conditioning may seem surprising, a few words of explanation may be in order. The initial decrease is due to the occurrence of the unreinforced CS_1-CS_2 pair in that the onset of the CS_1-CS_2 pair yields a positive Δx_1 that is followed by a negative Δy at the time of termination of CS_1 and CS_2. The negative Δy occurs because, with an unreinforced pair, no US onset occurs at the time of CS_1-CS_2 offset and thus there is nothing to cause the neuronal response to be sustained. The drive-reinforcement learning mechanism yields negative Δw's whenever a positive Δx_i is followed within τ time steps by a negative Δy. Thus, the excitatory weight associated with CS_1 decreases initially in stage two of conditioning. Similarly, the inhibitory weight associated with CS_2 is decreasing (i.e., becoming more negative or becoming stronger in terms of its absolute value) because CS_2 onset yields a positive Δx_2 that is followed by a negative Δy at the time of CS_1-CS_2 offset. The excitatory weight associated with CS_1 ceases to decrease and starts increasing when the conditioned inhibition becomes sufficient, such that the positive Δy following the onset of CS_1 unpaired with CS_2 is larger than the negative Δy following the onset of CS_1-CS_2 paired. The inhibitory weight associated with CS_2 continues to decrease (become more strongly inhibitory) because its onset, yielding a positive Δx_2, continues to be followed by a negative Δy until the conditioned inhibition of CS_2 becomes sufficient to cancel the conditioned excitation

of CS_1, at which point the CS_2 inhibitory weight, $w_2(I)$, approaches its asymptotic level. At the same time, the CS_1 excitatory weight, $w_1(E)$, approaches its asymptotic level, equal to its prior asymptotic level, because when the CS_2 conditioned inhibition cancels the CS_1 conditioned excitation, the reinforcement of CS_1 unpaired is the only event in each trial that yields a nonzero Δy following a positive Δx. Thus, toward the end of stage two conditioning, the situation in terms of positive Δx's followed by nonzero Δy's is similar to that which occurred in stage one.

Extinction and reacquisition effects

When conditioned excitation develops in conjunction with a CS, as was the case for CS_1 at the conclusion of stage one (trials 1-70) and stage two (trials 71-200) of conditioning in Figure 13, if the CS continues to be presented in a third stage of conditioning, this time without reinforcement, then Pavlov (1927) observed that the CR extinguishes; i.e., the CR decreases in magnitude, reaching zero with a sufficient number of unreinforced presentations of the CS. In addition, Pavlov inferred that conditioned inhibition developed during the extinction process because he observed "spontaneous recovery" of the CR with time and he also observed more rapid reacquisition of the CR if reinforced presentations of the CS were resumed. The predictions of the drive-reinforcement model are consistent with Pavlov's observations and inferences. Note that in stage three (trials 201-300) of conditioning in Figure 13, where CS_1 is presented without reinforcement, the CS_1 excitatory weight, $w_1(E)$, declines and the CS_1 inhibitory weight, $w_1(I)$,

grows stronger, until they cancel one another, at which time the CR will no longer appear.

Perhaps a few words are in order regarding the phenomenon of spontaneous recovery following extinction. Spontaneous recovery refers to the tendency of an extinguished conditioned response to return after the CS is not presented for some period of time. It seems that spontaneous recovery could be due to the state of the nervous system changing sufficiently with time so that the conditioned inhibition that may develop during the process of extinction becomes less effective. [As noted above, Pavlov (1927) believed that conditioned inhibition developed during the process of extinction. However, Rescorla (1969, p.87) has stated that "There is only meager evidence bearing on this question".] If the hypothesized conditioned inhibition were to become less effective because a change in the state of the nervous system resulted in fewer of the conditioned inhibitory synapses being active, then it would become easier for the conditioned response to manifest again. If this explanation of spontaneous recovery is correct, a neuronal model would not be expected to predict the phenomenon. A network model would be required to generate the prediction.

In the third stage of conditioning in Figure 13, the drive-reinforcement model makes one further prediction that has not yet been discussed. In this simulation, not only was CS_1 presented unreinforced in stage three but the CS_1-CS_2 pair was also presented unreinforced. Pavlov (1927) observed that under these circumstances, the conditioned excitation associated with CS_1 extinguished but the conditioned inhibition associated with CS_2 did not. This is predicted by

the drive-reinforcement model. In the third stage of conditioning in Figure 13, notice that the inhibitory weight, $w_2(I)$, remains unchanged during the unreinforced presentations of the CS_1-CS_2 pair. This prediction of the drive-reinforcement model differs from that of the Rescorla-Wagner model of classical conditioning. As Rescorla and Wagner (1972) point out, their model is inconsistent with the experimental evidence of conditioned inhibition studies in that the model predicts the extinction of conditioned inhibition. The drive-reinforcement model does not make this prediction because the positive Δx occurring at the time of CS_2 onset is not followed by a positive Δy.

Pavlov (1927) reported that after extinction of a CR, if reinforced presentations of the CS were resumed, then the CR would be reacquired more rapidly than during the first series of reinforced trials. The drive-reinforcement model predicts this reacquisition effect, as can be seen in Figure 14 where delay conditioning occurs in stage one (trials 1-70), extinction of the CR occurs in stage two (trials 71-140), and reacquisition of the CR occurs in stage three (trials 141-200). When measured to an accuracy of three significant figures, the CS excitatory weight reached its asymptotic level in 61 trials in stage one but only required 47 trials to reach the same level in stage two. This effect occurs because, during reacquisition, the CS excitatory weight begins at a higher level than during the initial acquisition process. It may be noted that this prediction of the drive-reinforcement model differs from that of the Rescorla-Wagner (1972) and Sutton-Barto (1981) models in that the latter two models do not predict the more rapid reacquisition of conditioned responses.

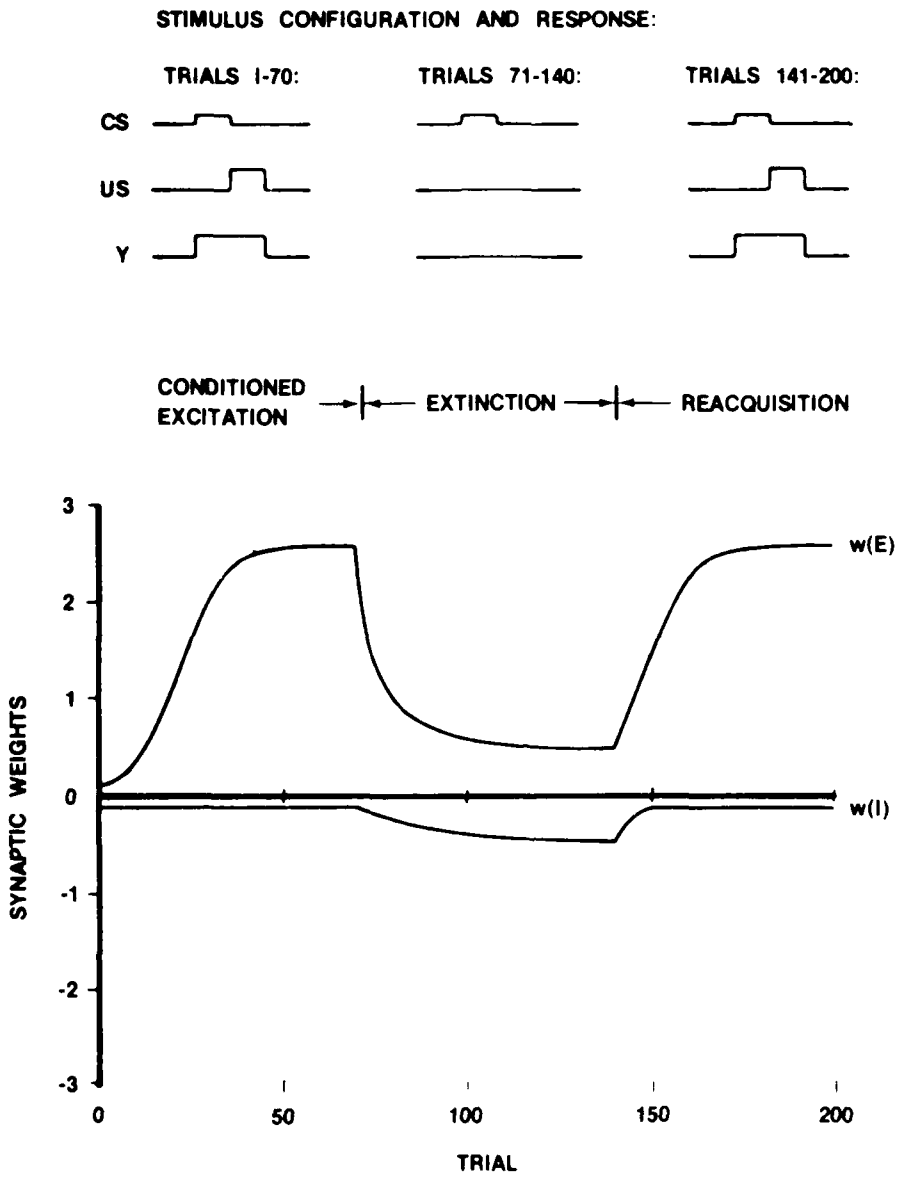


Figure 14. Results of a simulated three-stage classical conditioning experiment in which the drive-reinforcement model's predicted rate of reacquisition of a CR in stage 3 (trials 141-200) after extinction in stage 2 (trials 71-140) is compared with the predicted rate of initial acquisition in stage 1 (trials 1-70). Consistent with experimental evidence demonstrating that reacquisition occurs more rapidly, the drive-reinforcement model predicts that acquisition in stage 1 will require 61 trials as compared with 47 trials for reacquisition in stage 3. (See text and Appendix for details.)

Backward conditioning

In backward conditioning, the onset of the US precedes the onset of the CS. There have been conflicting reports regarding whether backward conditioning leads to conditioned excitation or conditioned inhibition (e.g., see review by Gormezano, Kehoe, and Marshall, 1983). Mahoney and Ayres (1976) sought to design experiments that would clarify some of the issues and they concluded that conditioned excitation did result from backward conditioning. At this time, the consensus appears to be that backward conditioning can lead to conditioned excitation initially but that extended backward conditioning usually yields conditioned inhibition (Pavlov, 1928; Rescorla, 1969; Wagner and Terry, 1975; Heth, 1976; Schwartz, 1984; Flaherty, 1985; Dolan, Shishimi, and Wagner, 1985). The initial conditioned excitation may be due to transient effects associated with global brain processes such as arousal triggered by the onset of the surprising US. In this view of backward conditioning, the hypothesized underlying process is one of conditioned inhibition which prevails with extended conditioning, after the US has come to be expected. The predictions of the drive-reinforcement model are consistent with this hypothesis, as can be seen in Figure 15. In Figure 15(a), forward conditioning is shown for a CS, the onset of which occurs two time steps before the onset of the US. In Figure 15(b), backward conditioning is shown for the same CS and US, in this case with the onset of the CS following the onset of the US by two time steps. The drive-reinforcement model predicts that backward conditioning will lead to conditioned inhibition, consistent with the experimental results obtained in most cases of extended backward conditioning. However, regarding these

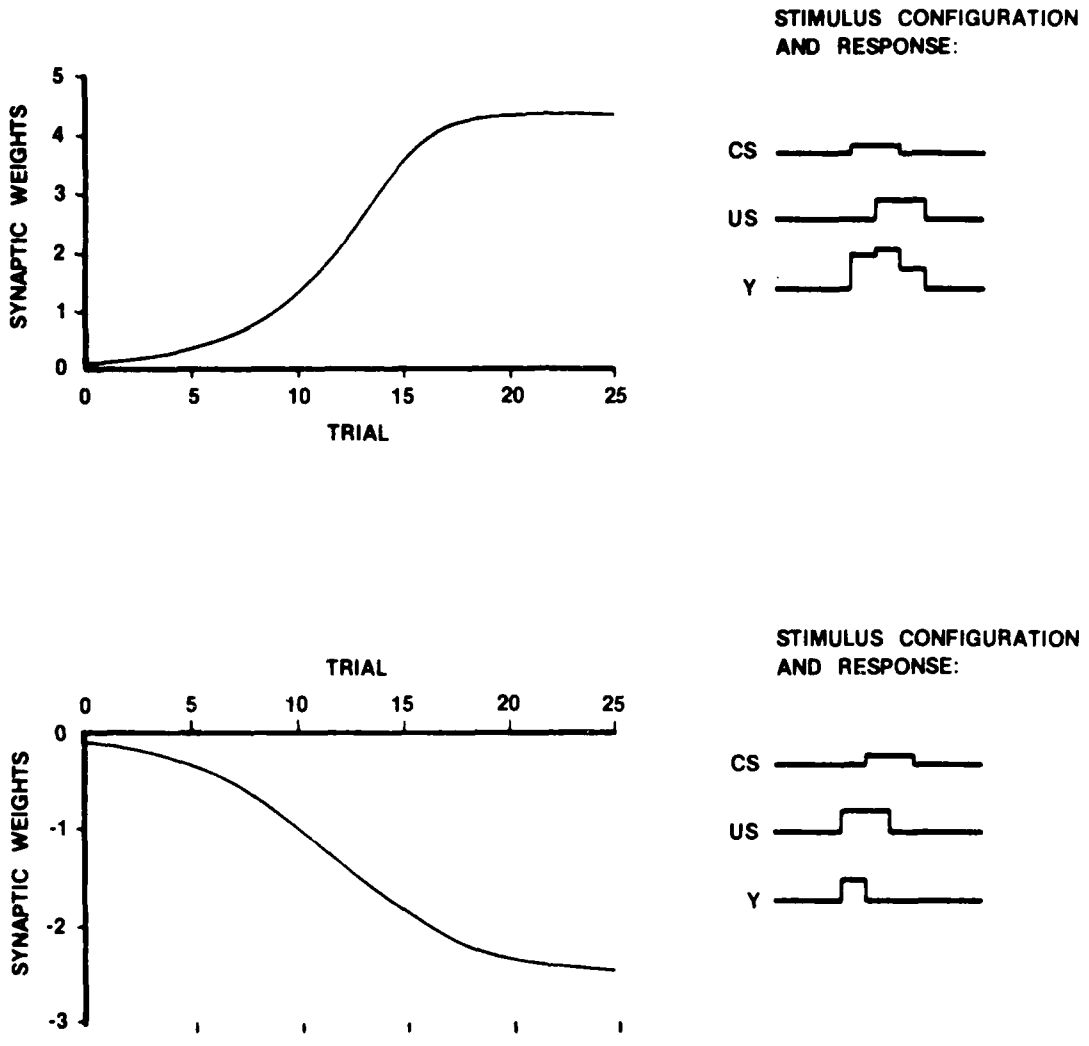


Figure 15. Results of simulated classical conditioning experiments in which the drive-reinforcement model's predictions for (a) forward and (b) backward conditioning are compared. Consistent with the experimental evidence, the model predicts that conditioned inhibition will result from backward conditioning, in contrast to conditioned excitation being predicted as the result of forward conditioning. (See text and Appendix for details.)

experimental results, J. W. Moore (personal communication, June 18, 1986) suggests that one caveat is in order: "... no studies have used the requisite combination of summation and retardation tests to assess the presumed learned inhibitory properties instilled by backward training."

Blocking and overshadowing

Temporal contiguity between a CS and US is fundamental to classical conditioning. This has long been understood to be the case. But while temporal contiguity is necessary, Kamin (1968, 1969) has demonstrated that it is not sufficient. Kamin has shown that a CS must also have predictive value. That is to say, there must be a contingent relationship between the CS and US as well as a relationship of temporal contiguity; otherwise, no conditioning will occur. Kamin demonstrated this by first reinforcing CS_1 with a US until an asymptotic level of associative strength was reached. Then he added CS_2 such that CS_2 was presented simultaneously with CS_1 and both were reinforced. Kamin showed that no or very little associative strength developed between CS_2 and the US. The first CS was said to have blocked conditioning of the second CS.

The drive-reinforcement model predicts the phenomenon of blocking, as can be seen in Figure 16. In this simulated blocking experiment, CS_1 is reinforced by the US in the first stage of conditioning (trials 1-100), until the CS_1 excitatory weight has approached its asymptotic level. Then, in stage two of conditioning (trials 101-160), CS_1 and CS_2 are presented simultaneously and reinforced with the US. It is seen that the CS_2 excitatory weight remains unchanged during the second stage of conditioning. Consistent with the experimental evidence, the

drive-reinforcement model predicts that conditioning of CS_2 will be blocked by CS_1, due to the previous conditioning of CS_1.

US preexposure effects may be due to the phenomenon of blocking (Mis and Moore, 1973). If an animal experiences a number of US presentations prior to experiencing paired presentations of a CS and the US, the result is that the conditioning process is retarded. This effect may be due to the experimental context, during US preexposure, becoming a blocker for subsequent conditioning [see review by Flaherty (1985) and, e.g., Balsam and Schwartz (1981)].

A question in animal learning theory has been whether contingency aspects of classical conditioning derive from limitations on the amount of associative strength available so that, in effect, stimuli must compete for the available associative strength (Rescorla and Wagner, 1972) or whether, in effect, stimuli must compete for an animal's attention (Sutherland and Mackintosh, 1971; Mackintosh, 1975; Moore and Stickney, 1980, 1985). The alternative hypotheses are not mutually exclusive. The drive-reinforcement neuronal model's predictions are consistent with the hypothesis that there are limitations on the associative strength available to stimuli. However, the neuronal model does not rule out the involvement of higher level attention mechanisms.

In the case of the drive-reinforcement model, it can be seen that the limits on $\Delta y(t)$ serve to limit the amount of associative strength available to competing stimuli. $y(t)$ is bounded such that it is less than or equal to $y'(t)$, the maximal frequency of firing of the neuron.

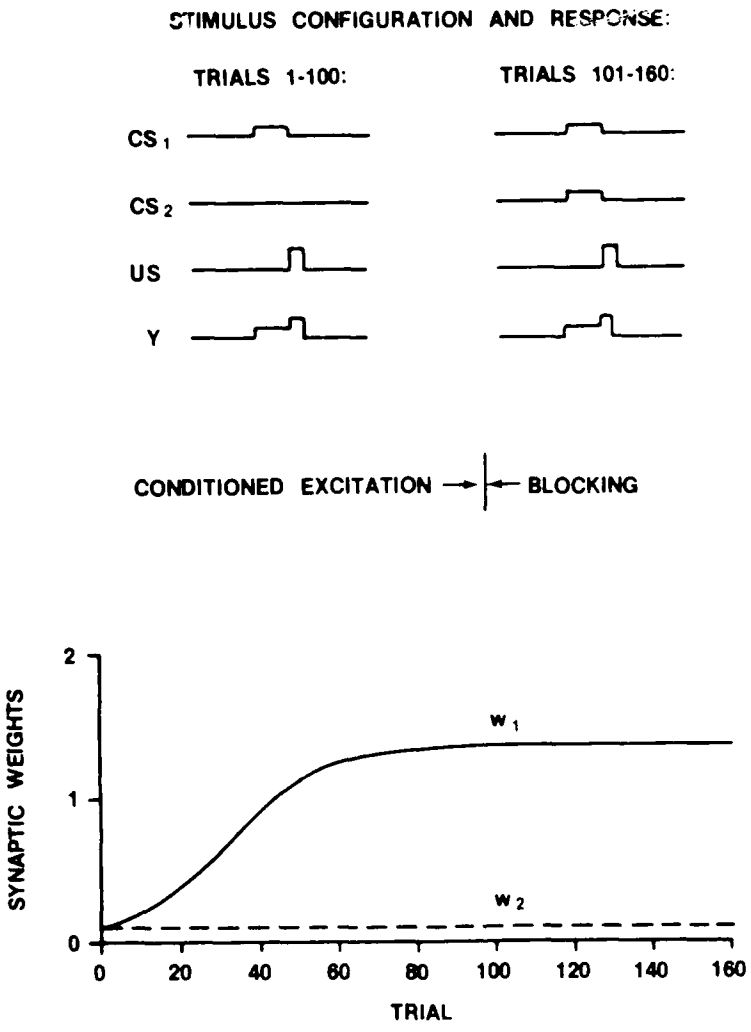


Figure 16. The drive-reinforcement model's predictions of the effects of a blocking stimulus. Consistent with the experimental evidence, the model predicts that after delay conditioning of CS_1 in stage 1 (trials 1-100), conditioning of CS_2, presented simultaneously with CS_1 in stage 2 (trials 101-160), will be blocked. The CS_2 excitatory synaptic weight, w_2, does not change in stage 2. (See text and Appendix for details.)

For nonoverlapping CSs and USs, the upper bound on $y(t)$ may actually be less that $y'(t)$ because, in this case, $y(t)$ never exceeds the amplitude of the UR. Thus, as was seen in Figure 16, if CS_1 has been reinforced until an asymptotic level of conditioning is reached, subsequent conditioning of a second stimulus, CS_2, will be blocked if the second stimulus forms a compound with the first and the onsets of CS_1 and CS_2 are simultaneous. What happens is that in stage 1 of conditioning, the positive $\Delta x_i(t-j)$ associated with CS_1 interacts with the subsequent positive $\Delta y(t)$ induced by the onset of the US, causing CR_1 to grow and thus diminishing $\Delta y(t)$ with each trial. Eventually the positive $\Delta y(t)$ associated with US onset diminishes to the point where its effect is cancelled by the subsequent effect of the negative Δy associated with US offset. The amplitude of CR_1 has grown to the point where there is no room for the generation of a net positive Δy subsequent to a positive Δx_i when CS_2 is introduced as part of a compound. Thus, consistent with the experimental evidence and consistent with the hypothesis of Rescorla and Wagner (1972) that there are limitations on the associative strength available to stimuli, the drive-reinforcement model predicts that conditioning will be blocked with respect to CS_2.

A variant of blocking is <u>overshadowing</u> (e.g., Baker, 1968; Courillon and Bitterman, 1982), first reported by Pavlov (1927), in which two or more simultaneous CSs are reinforced in a single stage of conditioning. In this type of experiment, it is observed that the more salient stimulus acquires the greatest associative strength, in effect, partially blocking

conditioning of the other stimuli. The drive-reinforcement model predicts overshadowing, as may be seen in Figure 17. In this simulated classical conditioning experiment, three simultaneous CSs are reinforced by a US. CS_1 and CS_2 are of equal amplitude. The amplitude of CS_3 is twice that of either of the other two CSs. Consistent with the experimental evidence, the drive-reinforcement model is seen to predict that the CS_3 excitatory weight will achieve a substantially higher asymptotic value than the equal and lower asymptotic values achieved by the CS_1 and CS_2 excitatory weights. This effect occurs with the drive-reinforcement model because the change in the presynaptic frequency of firing upon CS onset is greater for CS_3 than it is for CS_1 or CS_2. Thus, the CS_3 excitatory weight increases more rapidly, taking up a larger fraction of the total available associative strength than either the CS_1 or CS_2 excitatory weights.

Compound conditioning

In compound conditioning, multiple CSs are presented simultaneously or sequentially for reinforcement (or for nonreinforcement). Compound CSs have appeared in some of the simulated classical conditioning experiments discussed above, including those experiments involving conditioned inhibition, blocking, and overshadowing.

A compound conditioning experiment reported by Rescorla and Wagner (1972) can be utilized as a test of the drive-reinforcement model. The experimental results were obtained, Rescorla and Wagner (1972) note, in a previously unpublished study due to Wagner and Saavedra. The experiment involved comparing the effects of two CS-US configurations. In one

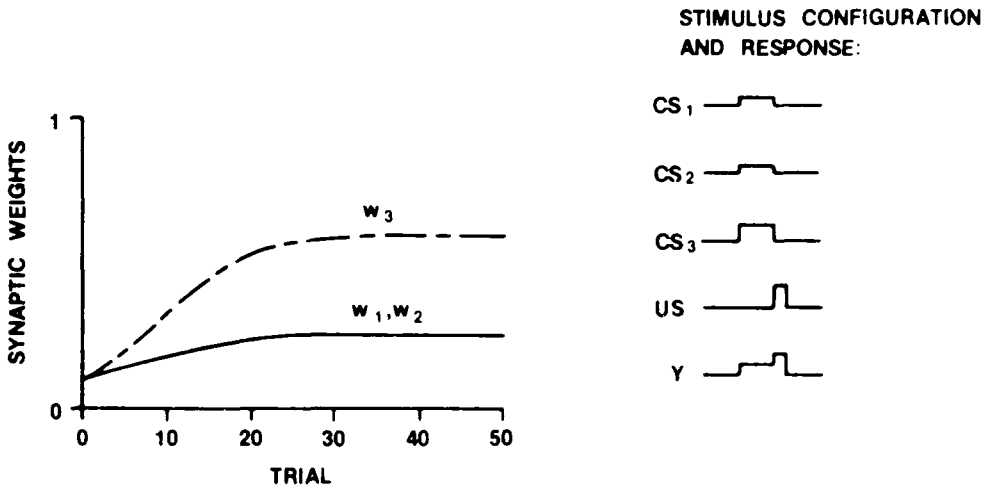


Figure 17. The drive-reinforcement model's predictions of the effects of stimulus salience on compound conditioning. Consistent with the experimental evidence, the model predicts that a more salient stimulus, CS_3, which has an amplitude of 0.4 will condition more rapidly and strongly than less salient stimuli, CS_1 and CS_2, each with an amplitude of 0.2. The asymptotic excitatory synaptic weight for CS_3 is more than double that of either the CS_1 or CS_2 asymptotic excitatory synaptic weights. Thus, the drive-reinforcement model predicts the phenomenon of overshadowing. (See text and Appendix for details.)

configuration, CS_1 occurring alone was reinforced and also CS_1 paired with CS_2 was reinforced. An example of such a CS-US configuration appears in Figure 18(a). In the second configuration, an example of which is shown in Figure 18(b), CS_1 occurring alone was not reinforced; only CS_1 paired with CS_2 was reinforced. In the case of the first configuration, where both CS_1 alone and CS_1-CS_2 paired were reinforced, the asymptotic associative strength of CS_1 was observed to be high and that of CS_2 was observed to be low. The ranking of the asymptotic associative strengths reversed when the second configuration was employed, in which CS_1 alone was not reinforced and CS_1-CS_2 paired was reinforced. These results are predicted by the drive-reinforcement model, as can be seen in Figure 18. In effect, what happens is that the CS that more reliably predicts the US comes to block the other CS.

Space limitations preclude the presentation of additional results of computer simulations of compound conditioning experiments. However, two other compound conditioning effects that are predicted by the drive-reinforcement model should be noted. In the case of the overexpectation paradigm, two stimuli, CS_1 and CS_2, are first individually conditioned to an asymptotic level, each stimulus being reinforced with the same US. Then, in a second stage of conditioning, the two stimuli are presented as a compound that is reinforced utilizing the same US as in the first stage. Rescorla and Wagner (1972) and Kremer (1987) report that the associative strengths of the two stimuli decrease in the second stage of conditioning. Furthermore, if an initially neutral stimulus, CS_3, is presented in compound with CS_1 and CS_2 in the

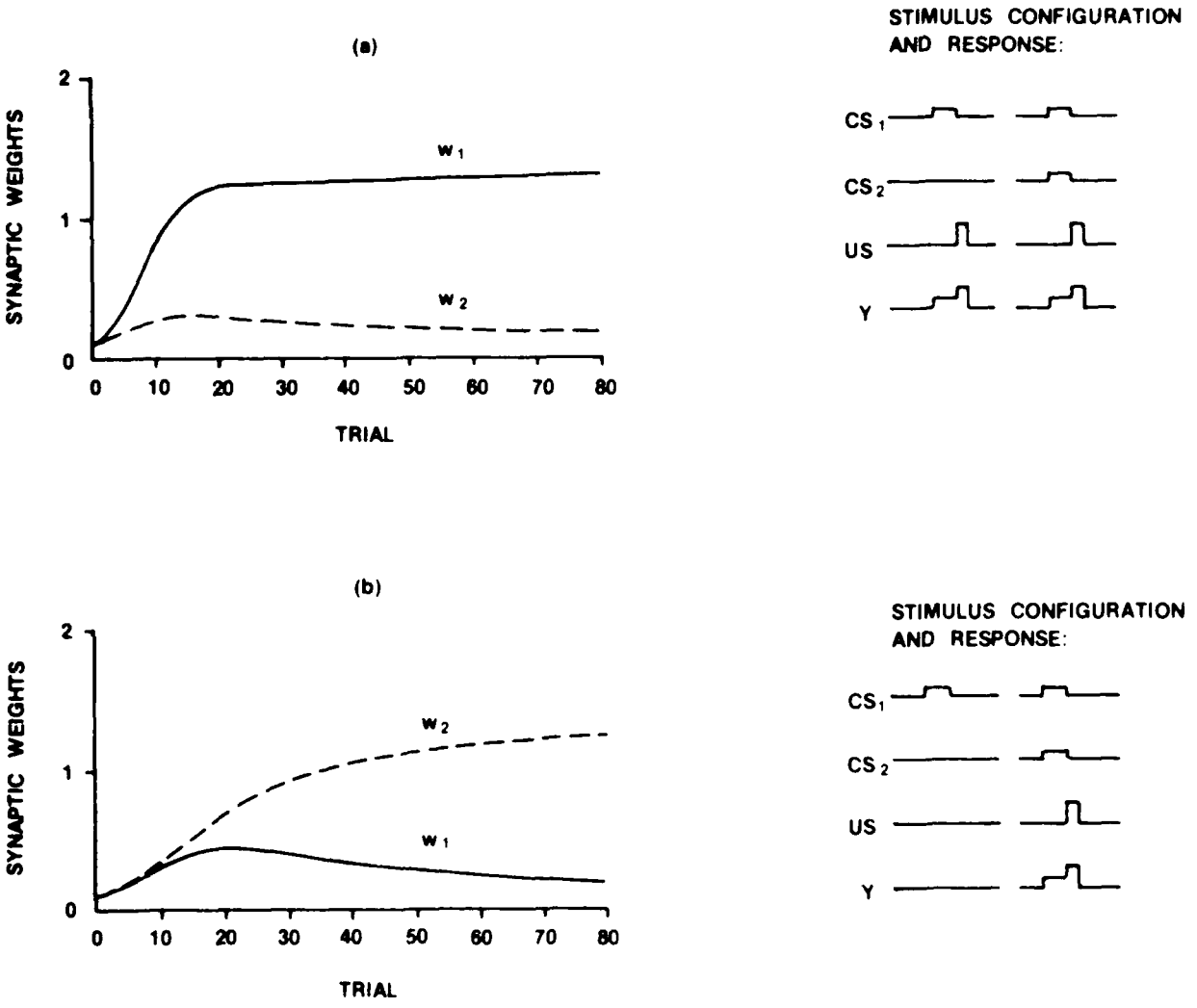


Figure 18. Results of simulated compound conditioning experiments in which the drive-reinforcement model's predictions for reinforced and nonreinforced CS's are compared. Consistent with the experimental evidence, in (a) the model predicts strong conditioning of CS_1 relative to CS_2, where both CS_1 alone and the CS_1-CS_2 pair are reinforced. Again consistent with the experimental evidence, in (b) the model predicts that the ranking of associative strengths for CS_1 and CS_2 will be reversed with respect to (a) when CS_1 alone is not reinforced and the CS_1-CS_2 pair is reinforced. (See text and Appendix for details.)

second stage of conditioning, CS_3 becomes a conditioned inhibitor. The drive-reinforcement model predicts these effects.

In the case of superconditioning, the compound to be reinforced consists of two stimuli, one initially neutral and the other a conditioned inhibitor by virtue of prior conditioning. Reinforcement of this compound is observed to yield an asymptotic associative strength for the initially neutral stimulus that is greater than the corresponding associative strength in a control experiment in which both stimuli are initially neutral (Rescorla, 1971; Wagner, 1971). The drive-reinforcement model predicts this effect.

Discriminative stimulus effects

The simulated classical conditioning experiments discussed above, the results of which were shown in Figure 18, involved compound conditioning and discrimination learning. Discrimination learning experiments test an animal's ability to discriminate between reinforced and nonreinforced CSs. A more complex example of a compound conditioning experiment that tests for discriminative stimulus effects is shown in Figure 19(a), where the compound CS_1-CS_3 is reinforced and the compound CS_2-CS_3 is not reinforced. For this CS-US configuration, experimental evidence reviewed by Rescorla and Wagner (1972) suggests that the asymptotic associative strengths will be high for CS_1, low for CS_3 and zero for CS_2. Actually, CS_2 is observed in the experiments to become a conditioned inhibitor. It is seen in Figure 19(a) that the drive-reinforcement model predicts these results. Furthermore, the drive-reinforcement model predicts that the combined associative

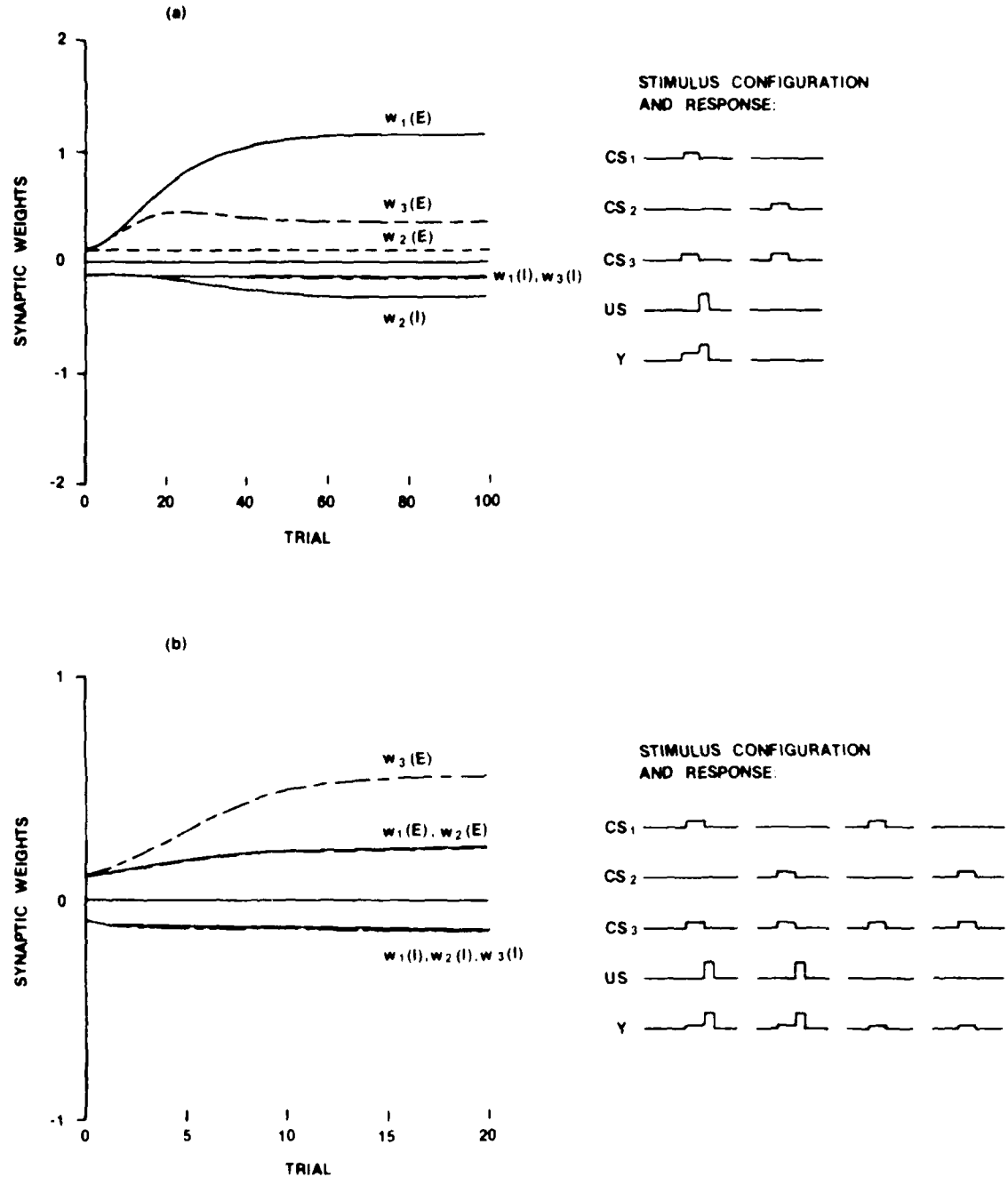


Figure 19. Results of simulated compound conditioning experiments in which the drive-reinforcement model's predictions of the effects of discriminative stimuli were determined for a more complex case than that portrayed in Figure 18. (See text and Appendix for details.)

strengths of CS_2 and CS_3 will increase initially and then decrease. This transient effect predicted by the model has been observed by experimentalists, as Rescorla and Wagner (1972) note.

A CS-US configuration similar to that shown in Figure 19(a) is shown in Figure 19(b). Rescorla and Wagner (1972) review the results of a study by Wagner, Logan, Haberlandt, and Price (1968) in which the discriminative stimulus effects of the CS-US configuration shown in Figure 19(a) were compared with the effects of the CS-US configuration shown in Figure 19(b). The CS-US configuration shown in Figure 19(b) represents a "pseudodiscrimination" experiment in that both compound CSs are reinforced sometimes and both are nonreinforced sometimes so it is actually a partial reinforcement experiment. Because of the similarity between the CS-US configurations in panels (a) and (b) of Figure 19, it is of interest to compare the experimental outcomes. It was found by Wagner et al. (1968) that while CS_3 was reinforced an equal fraction of the time in both the discrimination and the pseudodiscrimination training and occurred in compound with the same CSs, the eventual associative strength of CS_3, when tested alone, was much greater after pseudodiscrimination training than after discrimination training. This is predicted by the drive-reinforcement model, as can be seen by comparing the asymptotic synaptic weights for CS_3 in panels (a) and (b) of Figure 19. The net CS_3 asymptotic synaptic weight (i.e., the CS_3 asymptotic excitatory weight minus the absolute value of the CS_3 asymptotic inhibitory weight) in Figure 19(b) is approximately double that of the net CS_3 asymptotic synaptic weight in Figure 19(a). It

should be noted that the Rescorla-Wagner (1972) and Sutton-Barto (1981) models also correctly predict the experimental outcomes of the discrimination and pseudodiscrimination experiments just discussed, including the transient increase in the associative strength of the CS_2-CS_3 compound stimulus in the case of the discrimination training.

A variant of the drive-reinforcement neuronal model

The drive-reinforcement neuronal model, as specified above, requires that a positive change in presynaptic signal level occur in order that a synapse be rendered eligible for a change in its efficacy. It was noted earlier that the best argument for this constraint is that it yields a neuronal model that is consistent with the experimental evidence of classical conditioning. When the constraint is lifted so that $\Delta x_i(t-j)$ in equation (2) is allowed to assume any value, positive or negative, the neuronal model then frequently generates predictions that deviate substantially from the experimental evidence. An example is shown in Figure 20. The simulated classical conditioning experiment reported in Figure 20 is identical to one reported in Figure 16 (which was a blocking experiment) except that, in the case of Figure 20, $\Delta x_i(t-j)$ did not have to be greater than zero for the learning mechanism to be triggered. With this constraint removed, the CS_1 and CS_2 excitatory synaptic weights approach infinity because the negative Δx occurring at the time of CS offset is multiplied by the negative Δy occurring at the time of US offset. The behavior that is plotted in Figure 16 is a clear-cut presentation of the blocking phenomenon while that plotted in Figure

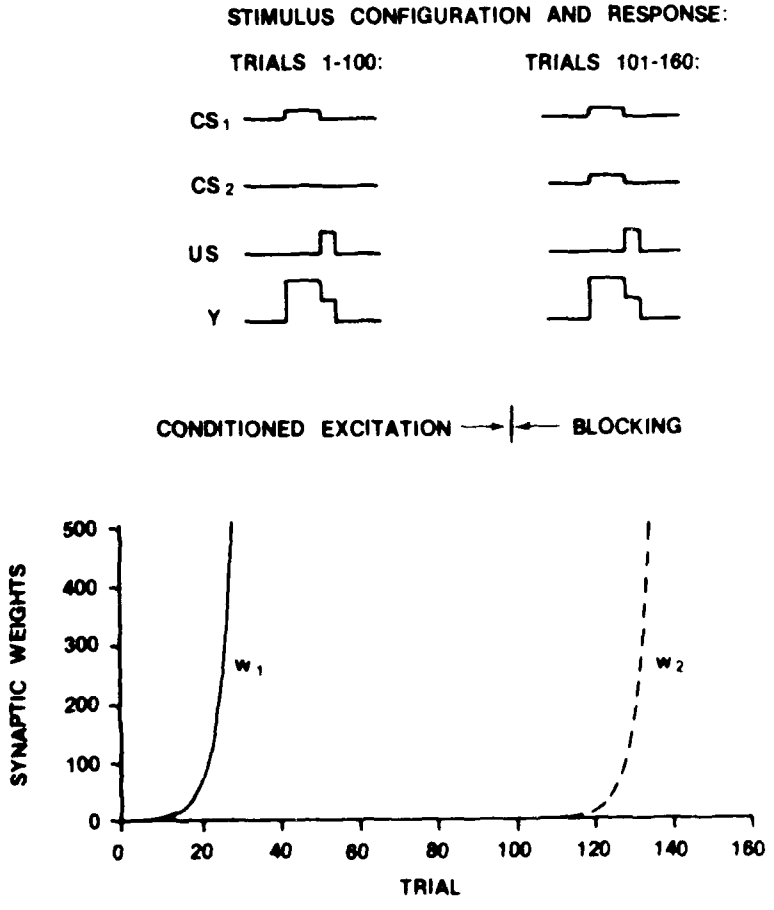


Figure 20. Results of a simulated blocking experiment that was identical to that reported in Figure 16 except that the drive-reinforcement model utilized to generate the predictions in Figure 16 rendered a synapse eligible for a change in its efficacy, w_i, only upon the occurrence of a positive change in the presynaptic signal level. To generate the predictions shown here, a variant of the drive-reinforcement model was employed, such that both positive and negative changes in presynaptic signal levels rendered a synapse eligible for a change in its efficacy. It is seen that the variant of the model employed here yields predictions that deviate markedly from the experimental evidence. Because these deviations are typical of this variant of the drive-reinforcement model, the other variant of the model, utilized to generate the predictions shown in Figures 4 through 19, seems more likely to reflect the function of biological neurons. (See text and Appendix for details.)

20 bears no discernable relationship to experimentally observed behavior in the case of blocking experiments.

Summary

By means of computer simulations of the drive-reinforcement neuronal model, it has been shown that the model correctly predicts classical conditioning phenomena in the following basic categories: delay and trace conditioning, conditioned and unconditioned stimulus duration and amplitude effects, partial reinforcement effects, interstimulus interval effects including simultaneous conditioning, second-order conditioning, conditioned inhibition, extinction, reacquisition effects, backward conditioning, blocking, overshadowing, compound conditioning, and discriminative stimulus effects.

SECTION 4

DRIVES AND REINFORCERS

The behavior of the proposed neuronal model may be understood in terms of two processes involving postulated neuronal drives and reinforcers. If weighted presynaptic signal levels are defined to be neuronal drives and weighted changes in presynaptic signal levels are defined to be neuronal reinforcers, then the drive-reinforcement learning mechanism operates such that neuronal drive induction promotes learned excitatory processes and neuronal drive reduction promotes learned inhibitory processes. The interplay between these two processes yields the classical conditioning phenomena discussed above.

In this section, definitions of drives and reinforcers at the level of the single neuron and at the level of the whole animal will be examined further. Then the relationship of the drive-reinforcement neuronal model to animal learning theory will be discussed. I will begin by offering precise definitions of drives and reinforcers, definitions motivated by the neuronal model as it may be viewed in the context of animal learning theory.

Definitions

For the drive-reinforcement neuronal model, neuronal drives are defined to be the weighted presynaptic signals, $w_i(t)\ x_i(t)$. These weighted presynaptic signals drive the neuron. Equation (1) is termed the drive equation because it specifies how neuronal drives, $w_i(t)\ x_i(t)$, are transformed into neuronal behavior, $y(t)$. Neuronal reinforcers are

defined to be the weighted changes in presynaptic signal levels, $w_i(t)$ $\Delta x_i(t)$. Neuronal reinforcement results from the net effect of all of the weighted Δx_i's experienced by a neuron at time, t. Neuronal reinforcers thus manifest as $\Delta y(t)$ and neuronal reinforcement is defined to be equal to $\Delta y(t)$. Note the distinction here: A neuronal reinforcer is a weighted change in signal level that the neuron experiences at a single synapse; neuronal reinforcement is defined to be the collective effect of the neuronal reinforcers, manifesting as the change in output, $\Delta y(t)$. Incremental neuronal reinforcement is defined to be an increase in the postsynaptic frequency of firing and decremental neuronal reinforcement is defined to be a decrease in the postsynaptic frequency of firing, with both increases and decreases in firing frequency measured over intervals not exceeding a few seconds.

In the drive-reinforcement neuronal model, changes in presynaptic signal levels play two roles. A change in presynaptic signal level, $\Delta x_i(t)$, renders the i^{th} synapse eligible for future reinforcement. The synaptic weight, $w_i(t)$, for such an eligible synapse changes if a subsequent change in postsynaptic signal level, Δy, occurs not more than τ time steps in the future. The other role for $\Delta x_i(t)$, when weighted by $w_i(t)$, is to contribute to (i.e., partially or wholly cause) $\Delta y(t)$ and thus reinforce synapses rendered eligible by earlier changes in presynaptic signal levels. In effect, $\Delta x_i(t)$ looks to the future with regard to its role in rendering a synapse eligible for reinforcement and to the past in contributing to the reinforcement of synapses rendered eligible earlier. Equation (2), the neuronal learning

mechanism, is termed the reinforcement equation because it specifies how neuronal reinforcers [$w_i(t)\ \Delta x_i(t)$'s manifesting collectively as $\Delta y(t)$] are transformed into changes in behavior [due to $\Delta w_i(t)$'s]. Thus, we see that equation (1), the drive equation, involves the processing of signal levels to yield behavior and equation (2), the reinforcement equation, involves the processing of changes in signal levels to yield learning.

It was noted earlier that the drive-reinforcement learning mechanism moves the onsets and offsets of pulse trains to earlier points in time. It should also be noted that, in doing this, the learning mechanism sets up the possibility of a chain of reinforcing events. Because of the way Δx's and Δy's interact in the model to yield Δw's, Δy's come to occur earlier in time, making them available to reinforce even earlier Δx's. Thus, chains of reinforcing Δx's and Δy's can be established beginning with a primary reinforcer (which will be defined below).

While the drive-reinforcement neuronal model appears complex relative to, say, the Hebbian model, this seems appropriate because the single neuron is coming to be recognized as a highly sophisticated cell. None of the operations proposed here seem incompatible with the known capabilities of the single neuron (e.g., see Woody, 1982, 1986).

The terms I have defined at a neuronal level mirror terms animal learning researchers have defined at the level of the whole animal. Additional terms may be defined in this way. For example, innate or primary neuronal drives may be distinguished from acquired neuronal drives. Primary neuronal drives are defined to have fixed synaptic

weights. Acquired neuronal drives are defined to have variable synaptic weights, under the control of the neuronal learning mechanism. Primary neuronal drives will include deficit related signals having an internal source (drives to eat and drink are examples) and unconditioned stimuli having an external source (food and water are examples). Acquired neuronal drives, likewise, are expected to have internal sources (as the result of possible conditioning, for example, of the hypothalamic reward and punishment centers) and to have external sources in the case of what becomes conditioned stimuli. The notion of acquired drives was first suggested by Miller (1951).

Psychologists have generally defined drives to include only the category of deficit related internal signals. I am broadening the definition to include any signal that drives a neuron. My definition of primary drives comes closer to the conventional definition of drives but, in this case, I still include (external) unconditioned stimuli as well as (internal) deficit-related signals. My point in changing the definition is to suggest that drives, defined in this broader fashion and at a neuronal level, can serve as the basis for a simpler and more rigorous learning theory.

I have noted that neuronal drives can have internal and external sources and can be primary (innate) or acquired. The same is true of neuronal reinforcers as they have been defined above. Unconditioned stimuli, for example, function as primary drives, yielding unconditioned responses. Unconditioned stimuli also function as primary reinforcers, yielding conditioned responses. The drive-reinforcement model suggests

that when an unconditioned stimulus is functioning as a neuronal drive, it is the signal level, itself, that is important [see equation (1)] and when an unconditioned stimulus is functioning as a neuronal reinforcer, it is the onset and offset of the signal that is important [see equation (2)].

I have defined drives and reinforcers in a straightforward fashion at a neuronal level. However, such clear-cut definitions have not proved to be possible at the level of the whole animal. For example, Toates (1985, p. 963) remarks that the notion of drive "has been around for a long time. No one seems to know quite why we need the concept, but we keep putting it on display. It tends therefore to assume a variety of uncertain functions." I am going to argue that we should not be surprised by this state of affairs. In the history of animal learning research, it has not been unusual for the notions of drives and reinforcers to be seen as problematic. When such notions are invoked at the level of the whole animal, this may be understandable. If the notions of drive and reinforcement are relatively straightforward at the level of the single neuron, as I am suggesting here, then we should not necessarily expect such notions to also be straightforward at higher levels. If neurons are classically conditionable cells in their own right, as the drive-reinforcement model suggests, then when millions or billions of such neurons interact in phylogenetically advanced nervous systems, the interactions may not be simple. That we can make as much sense out of whole brain function as we have, thanks to the dedication of animal learning researchers and many others, might even be seen as

surprising, considering the complexity of the neural network of, say, a dog. That Pavlov (1927) and those who worked with and after him were able to see their way through to a relatively clear view of classical conditioning suggests that brain function may not be as complex as we might have expected. However, as Gray (1975) demonstrates in an especially careful and incisive analysis, complications arise with the notions of drives and reinforcers at the level of the whole animal.

If the notion of drive has been problematic at the level of the whole animal, what about the notions of drive reduction and drive induction, postulated to function as reinforcers (e.g., see Mowrer, 1960)? I have suggested that, at a neuronal level, drive reduction and induction have straightforward roles to play with respect to the process of reinforcement. Assuming for the moment that the hypothesized drive-reinforcement neuronal model is correct, how might we expect neuronal drive reduction and induction to map onto the level of the whole animal? Let us consider an example. The global reward or "pleasure" centers discovered by Olds and Milner (1954) are known to be inhibitory (Fuxe, 1965) so they would be expected to yield <u>decremental</u> neuronal reinforcement. However, we know that the salivary reflex is <u>excited</u> by the taste of food. Also, the brain's global reward centers are presumably <u>excited</u> by the taste of food but they will, in turn, deliver <u>inhibition</u> throughout the nervous system. This inhibition, in some cases, is likely to reach inhibitory interneurons and, thus, in effect, could be translated into <u>excitation</u>. Disinhibition is known to play a major role in the nervous system (Roberts, 1980). We can see then that

there will be no clear-cut, simple mapping of excitation and inhibition into drives. Neither should we expect increases and decreases in excitation and inhibition (neuronal drive reduction and induction) to map in a clear-cut, simple way into global reinforcement (i.e., reward and punishment). In each case, the involved neural network will have to be considered before any mapping of neuronal drives and reinforcers into global drives and reinforcers can be established.

Evidence for this kind of complexity has been obtained by Keene (1973). Olds (1977, p. 95) has summarized Keene's findings as follows: "A family of neurons excited by aversive brain shocks and inhibited by rewarding ones was identified in the intralaminar system of the thalamus; and a second family accelerated by rewards and decelerated by punishments was observed with probes in the preoptic area." Keene's results demonstrate that the brain's global processes of reward and punishment can have opposite effects in different parts of the nervous system. Thus, the practical complexity of this situation at the level of the whole animal, reflecting perhaps the pragmatic decisions of the evolutionary process, may account for the problematic history of the psychological notions of drives and reinforcers.

Relationship of the drive-reinforcement neuronal model to animal learning theory

Having defined drives, reinforcers, and related terms at a neuronal level, and having acknowledged the complexities that arise around these concepts at the level of the whole animal, I will now discuss how the

drive-reinforcement neuronal model relates to theories of animal learning.

In this century, the study of learning began with stimulus-response (S-R) association psychology (Thorndike, 1911; Pavlov, 1927; Guthrie, 1935). In place of S-R association psychology, the drive-reinforcement neuronal model suggests what could be called ΔS- ΔR association psychology. The neuronal model suggests that it is not stimuli and responses that are associated but, rather, changes in stimuli and changes in responses except, of course, in the theoretical model I am proposing, it is neuronal ΔS's and ΔR's that are associated, not the ΔS's and ΔR's of the whole animal. At the level of the whole animal, we can expect a more complicated situation, as I have already discussed.

Hull (1943) confronted the complexities that arise at the level of the whole animal. As Hilgard and Bower (1975) note, Hull, in his herculean effort to systematize learning theory, was moving psychology from an S-R formulation to an S-O-R formulation, where "O" represented the state of the organism. Central to Hull's (1943) theory of learning was the definition of reinforcement as drive reduction. Hull (1952) went on to revise his position, concluding that reinforcement should be defined as drive-stimulus reduction. Actually, Hilgard and Bower (1975, p. 167) observe that "While favoring drive-stimulus reduction, Hull left the matter somewhat open, having vacillated between drive reduction and drive-stimulus reduction as essential to reinforcement" (Hull, 1952, p. 153). The drive-reinforcement neuronal model suggests that Hull may have been right on both counts; both drive reduction and drive-stimulus

reduction may function as reinforcers because both can result in Δy's. Thus, at a neuronal level, the distinction between drive reduction and drive-stimulus reduction appears to dissolve. We see a reason why drives should probably be defined more broadly than Hull considered.

Hull's narrower definition of drives resulted in another problem for his theory. Hull's identification of drives and drive reduction with physiological needs or tissue deficits did not seem to leave room for such phenomena as animal play and the learning that results. Mishkin and Petri (1984, p. 292) point out that "Shortly after Hull developed [his] ideas, a number of studies on curiosity, manipulation and exploration suggested that other motives, not obviously related to physiological needs, also generated learning." Mishkin and Petri go on to say that "The recognition that there are motives that have no apparent basis in tissue deficits or other physiological needs was one major factor that eventually led to the demise of the drive reduction theory of learning (Bolles, 1967)." The drive-reinforcement neuronal model solves the problems encountered with Hull's theories by moving from the level of the whole animal to the level of the single neuron, by suggesting a broader definition of drives, by allowing both drive reduction and drive induction to be reinforcing [consistent with Mowrer (1960)], and by not necessarily identifying drive reduction with reward.

Regarding the relationship of drive reduction to reward, Gray (1975) discusses the question of whether rewards and punishments should be associated with drive decrements and increments, respectively. Based on Gray's analyses and those of others whom he cites, I have come to the

conclusion that too close an identification of drive reduction with reward may not be warranted. The Darwinian process may have been more flexible in its approach as it evolved nervous systems. Therefore, I will not, in the theoretical framework I am proposing in this report, identify drive decrements with reward and drive increments with punishment even though, as generalizations, such identifications may be valid. There is nothing in the theoretical framework that requires such a rigid identification to make the theory workable.

After Hull, animal learning theory's next major step forward was due, in my opinion, to Mowrer (1960). A colleague of Hull's at Yale, Mowrer moved from Hull's drive reduction (or drive-stimulus reduction) theory to a symmetric theory in which learning was attributed to both drive reduction and drive induction. Also, in Mowrer's theory, classical conditioning was accepted as basic. Mowrer's emphasis on classical conditioning and on symmetric processes in learning has had a strong influence on the theoretical framework I am proposing in this report.

Since Mowrer proposed his theory, substantial theoretical and experimental advances have occurred in the understanding of classical conditioning phenomena. Model systems such as the rabbit nictitating membrane response are providing a refined understanding of classical conditioning at psychological and neurobiological levels (e.g., see Gormezano, 1972; Moore and Gormezano, 1977; Moore, 1979; Gormezano, Kehoe, and Marshall, 1983; Thompson, 1976; Thompson, Berger, and Madden, 1983). Also, the investigations of Kamin (1968) and Rescorla and Wagner

(1972) have clearly demonstrated contingency aspects of classical conditioning as distinguished from contiguity aspects.

Along with an increased understanding of classical conditioning has come a growing conviction on the part of some animal learning theorists that classical conditioning phenomena are fundamental to animal learning; instrumental conditioning phenomena are then de-emphasized by these theorists. Mowrer (1960) early on and Bindra (1976, 1978) more recently have been leaders in this movement. The drive-reinforcement neuronal model is consistent with this view. If brains are, fundamentally, classically conditionable systems, then this may be because they are composed of classically conditionable neurons, as the drive-reinforcement model suggests. Instrumental conditioning phenomena are then seen to arise out of a neural substrate that utilizes classical conditioning mechanisms. As Bindra (1976, p. 245) has noted: "Once it is explicitly assumed that the production of any specific instrumental response or of some of its act components is linked to one or more particular eliciting stimulus configurations, then the way becomes clear for interpreting instrumental learning in terms of the learning of stimulus-stimulus contingencies alone. The problem of instrumental training then becomes one of making certain response-eliciting stimuli highly potent motivationally, and this can be done through stimulus-stimulus contingency learning between the response-eliciting stimulus and the incentive stimulus." Research on autoshaping in which animals shape their behavior without a response-reinforcer contingency supports this position (Brown and Jenkins, 1968; Jenkins and Moore, 1973). As

expressed by Flaherty, Hamilton, Gandelman, and Spear (1977, page 243), "the law of effect is apparently not necessary for the development of instrumental-like behavior."

Another way of viewing Bindra's theoretical position is as part of a movement away from drive reduction theories that emphasize internal deficit signals and toward incentive-motivation theories (Bindra, 1968; Bolles, 1972). Incentive-motivation theories suggest that "motivated behavior results not only from the 'push' of internal, deficit signals but also from the 'pull' of external, incentive stimuli" (Mogenson and Phillips, 1976, p. 200, emphasis is that of the quoted authors). It may be noted that neuronal drives, as defined earlier in this report, include both internal deficit signals and external incentive stimuli.

While finding myself in sympathy with those who emphasize that classical conditioning is fundamental to learning, I do not want to go too far in that direction. Miller and Balaz (1981) note that classical conditioning has often been seen as involving the learning of stimulus-stimulus associations while instrumental conditioning has often been seen as involving the learning of stimulus-response associations or, in the case of Mackintosh (1974), response-reinforcement associations. Frequently animal learning theorists have chosen one particular class of associations as being fundamental and then have ruled out other classes of associations. Bindra (1976, 1978), for example, suggests that learning does not have to do with the forming of stimulus-response associations but rather with the learning of contingencies between stimuli. This question of which class of associations is fundamental to

learning has been debated by animal learning theorists for decades. The drive-reinforcement neuronal model suggests that it may not be necessary to choose one class of associations over another. Solomon (1981, p. 2) observes: "One persisting question is 'what is learned?' The four candidates from the past were S-S associations, S-R associations, R-reinforcer associations and S-reinforcer associations." Solomon goes on to say: "It appears ... that associations of all four kinds can be established with the right procedures." The drive-reinforcement model allows for all four possibilities, suggesting that any of the four classes of associations will form when neuronal signals representing stimuli, responses, and reinforcers occur in appropriate temporal relationships. If a stimulus, response, or reinforcer results in a positive Δx_j that is followed within the interval, τ, by another stimulus, response, or reinforcer that yields a Δy at the same neuron, then an association will form. Thus, an implication of the drive-reinforcement model is that, at a neuronal level, classical conditioning, instrumental conditioning, drive-reduction and induction, response-reinforcement, and incentive-motivation theories may all describe associations that can form in the nervous system. However, it is not the presence of signals representing stimuli, reponses, or reinforcers that is required, according to the drive-reinforcement model, but rather changes in signal levels representing the onsets and offsets of stimuli, reponses, and reinforcers.

A drive-reinforcement theory of learning

What kind of theory of learning is implied then by the drive-reinforcement neuronal model? At this point, I will sketch one possible form such a theory might take.

Three principles would appear to be fundamental to what I will call a drive-reinforcement theory of learning:

(1) Primary neuronal drives are the foundation upon which all learning rests.

(2) Neuronal reinforcers are changes in neuronal drive levels. Neuronal drive induction promotes learned excitatory processes. Neuronal drive reduction promotes learned inhibitory processes. Together, these processes yield acquired drives or learning.

(3) The neuronal learning mechanism correlates earlier changes in presynaptic signals with later changes in postsynaptic signals yielding changes in the efficacy of synapses. A change in the efficacy of a synapse is proportional to the current efficacy of the synapse.

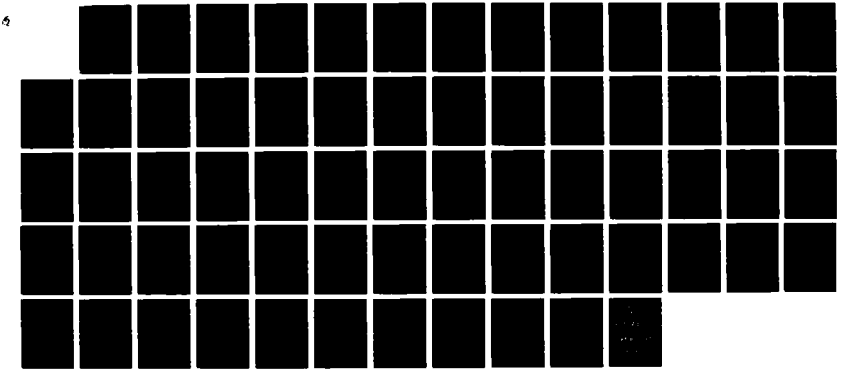
If these principles should turn out to be correct at a neuronal level, how should we expect such mechanisms to manifest at the level of the whole animal or what I will call the network level? Neuronal drives might be expected to emerge at the network level as the positive and negative feedback loops that control behavior. As examples, consider a blood glucose detector that provides an internal primary drive signal (this is what animal learning psychologists have customarily referred to

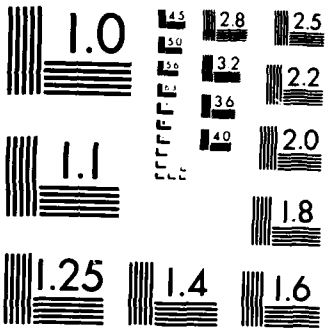
as a drive) or the taste of food that provides an external primary drive signal (what animal learning psychologists have customarily referred to as an unconditioned stimulus). These primary drive signals are parts of innate negative feedback loops that are associated with what are termed the hunger drive and the salivation reflex. These feedback loops cause the blood glucose level to rise because the animal is driven to eat and assist in causing food to disappear from the mouth and be digested because the animal is driven to salivate. More generally, feedback loops representing drives include mating behavior, drinking behavior, behaviors associated with the approach to and consumption of prey, and behaviors associated with the attack of or flight from predators. In general, behaviors can be classified as approach or avoidance (Mowrer, 1960). We might expect approach behavior to be supported by positive feedback loops and avoidance behavior to be supported by negative feedback loops. Positive and negative feedback loops that emerge at the level of the whole animal will be defined to be network drives, as distinguished from the neuronal drives defined earlier. Neuronal drives may be the more atomistic basis of network drives.

Primary network drives are the innate goal
Acquired network drives are the learned goal
basis of the hypothesized drive-rei
mechanism, it is expected that acquir
constructed on top of the prim
acquired) drive levels vary,
reinforcement and this

AD-A188 378 A NEURONAL MODEL OF CLASSICAL CONDITIONING(U) AIR FORCE 2/2
WRIGHT AERONAUTICAL LABS WRIGHT-PATTERSON AFB OH
A H KLOPF OCT 87 AFWAL-TR-87-1139

UNCLASSIFIED F/G 12/9 NL





positive and negative feedback loops). In this way, the process of learning is hypothesized to be sustained, with drives being built on top of drives. (Actually, in some cases, the process will not involve the acquisition of new drives so much as it will the refinement of current drives.) When acquired network drives become sufficiently complex, cognitive phenomena may begin to emerge.

To support the process of drive acquisition or learning at the network level, global centers that can broadcast generalized "start" and "stop" signals may be helpful. Such signals could serve to introduce appropriate Δy's in the network at crucial times, thus rendering the overall activity of the network coherent. Such may be the roles of the global reward and punishment centers discovered, respectively, by Olds and Milner (1954) and by Delgado, Roberts and Miller (1954). Consistent with this idea, global reward centers appear to employ inhibitory neurotransmitters (Stein, Wise, and Belluzzi, 1977) that may function as "stop" signals and global punishment centers appear to employ excitatory neurotransmitters (Stein, Wise, and Belluzzi, 1977) that may function as "start" signals. That a reward center should generate "stop" signals might seem paradoxical with respect to some behaviors, but disinhibitory mechanisms that are prevalent in the nervous system (Roberts, 1980) may make such an approach workable by enabling releasing mechanisms to be implemented where necessary. It should also be noted that if reinforcers are changes in drive levels, then global drive and reinforcement centers can be one and the same. A center's output will constitute a drive and a change in a center's output will constitute a

as a drive) or the taste of food that provides an external primary drive signal (what animal learning psychologists have customarily referred to as an unconditioned stimulus). These primary drive signals are parts of innate negative feedback loops that are associated with what are termed the hunger drive and the salivation reflex. These feedback loops cause the blood glucose level to rise because the animal is driven to eat and assist in causing food to disappear from the mouth and be digested because the animal is driven to salivate. More generally, feedback loops representing drives include mating behavior, drinking behavior, behaviors associated with the approach to and consumption of prey, and behaviors associated with the attack of or flight from predators. In general, behaviors can be classified as approach or avoidance (Mowrer, 1960). We might expect approach behavior to be supported by positive feedback loops and avoidance behavior to be supported by negative feedback loops. Positive and negative feedback loops that emerge at the level of the whole animal will be defined to be network drives, as distinguished from the neuronal drives defined earlier. Neuronal drives may be seen as the more atomistic basis of network drives.

Primary network drives are the innate goals of the organism. Acquired network drives are the learned goals of the organism. On the basis of the hypothesized drive-reinforcement neuronal learning mechanism, it is expected that acquired network drives are, in effect, constructed on top of the primary network drives. When primary (and acquired) drive levels vary, these variations in drive levels constitute reinforcement and this reinforcement will spawn new drives (acquired

reinforcer. Consistent with this theoretical possibility, drive and reinforcement centers in the limbic system and hypothalamus appear to be so close together (Olds, 1977) as to be, perhaps, identical.

SECTION 5

EXPERIMENTAL TESTS

In the computer simulations reported above, the drive-reinforcement neuronal model has been demonstrated to be consistent, in general, with the experimental evidence of classical conditioning. However, such a demonstration involves comparing theoretical predictions of a neuronal model with experimental evidence obtained from whole animals. To some extent, whole animal data has to be problematic vis a vis the predictions of a neuronal model. The effects of multiple interacting neurons , the effects of the brain's many interacting subsystems and, in general, the effects of the global architecture of the brain will, of course, influence whole animal data. All of these effects, collectively, I will refer to as network effects to distinguish them from neuronal (meaning single neuron) effects. Network effects will preclude rigorous experimental tests of any neuronal model in terms of whole animal data. Tests at a neurobiological level will be required. Fortunately, such experimental tests are becoming feasible and, indeed, results to date encourage the notion that classical conditioning phenomena may manifest at the level of the single neuron, as the drive-reinforcement model suggests. [See reviews by Kandel and Spencer (1968), Mpitsos, Collins, and McClellan (1978), Thompson, Berger, and Madden (1983), Farley and Alkon (1985), Woody (1986), Carew and Sahley (1986), and Byrne (1987). See also Hawkins and Kandel (1984) and Kelso and Brown (1986).]

Instrumental conditioning experiments at the level of the single neuron are also becoming feasible (Stein and Belluzzi, in press).

At this point, perhaps a note is in order regarding the semantics I am adopting. When I suggest that a single neuron may manifest classical conditioning phenomena, the "single neuron" I am referring to includes the synapses that impinge upon it. Those synapses, of course, come from other neurons or from sensory receptors and, in that sense, what I am referring to as a phenomenon involving a "single neuron" is, in fact, a multineuron or neuron and receptor phenomenon. The point, though, is that a single neuron may be undergoing the conditioning, as distinguished from alternative theoretical models that can be envisioned in which whole circuits consisting of many neurons would be the lowest level at which conditioning could occur. An implication of the drive-reinforcement neuronal model is that classical conditioning is not an emergent phenomenon but, rather, that the ability to undergo classical conditioning is a fundamental property of single cells.

Actually, the hypothesized drive-reinforcement learning mechanism could be implemented at a lower level than that of the single neuron. Minimally, what would seem to be required would be two synapses interacting such that one synapse would deliver the signal corresponding to $\Delta x_i(t-j)$, reflecting the onset of the CS, and the other synapse would deliver the signal corresponding to $\Delta y(t)$, reflecting the onset or offset of the US. Evidence of such interactions between synapses has been obtained in investigations of classical conditioning in Aplysia. The learning mechanism appears to involve what is termed

activity-dependent amplification of presynaptic facilitation (Hawkins, Abrams, Carew, and Kandel, 1983) or activity-dependent neuromodulation (Walters and Byrne, 1983) of sensory neuron terminals. The optimal interstimulus interval between activation of the sensory neuron terminal representing the CS and activation of the facilitator neuron terminal representing the US has been found to be about 500 ms (Carew, Walters, and Kandel, 1981; Hawkins, Carew, and Kandel, 1986). While the evidence for conditioning at a neuronal level in Aplysia has been interpreted as suggesting a presynaptic learning mechanism, Farley and Alkon (1985) indicate that the sites of the changes may not be exclusively presynaptic.

Whether presynaptic or postsynaptic processes or both underlie learning is a question that has been investigated theoretically (Zipser, 1986) and experimentally (Carew, Hawkins, Abrams, and Kandel, 1984). In this report, I have formulated the drive-reinforcement learning mechanism in terms of postsynaptic processes although,as discussed above, the learning mechanism could be implemented in an exclusively presynaptic form. Apart from activity-dependent amplification of presynaptic facilitation or activity-dependent neuromodulation offering a possible implementation of the drive-reinforcement learning mechanism, other possibilities can be envisioned that would still involve less than a whole neuron. Portions of dendritic trees and their impinging synapses might function in a manner analogous to the model I have envisioned for the whole neuron. Thus there are a range of possibilities for implementation of the drive-reinforcement learning mechanism, extending

from what is perhaps a minimal two-synapse interaction on the low end ranging through portions of dendritic trees functioning as a basic unit of learning, up through the level at which a single neuron functions as the basic unit and beyond to the point where the whole organism is treated as a single unit. Variations of the drive-reinforcement model may have relevance at each of these levels, even though the learning mechanism seems to lend itself naturally to implementation at a neuronal level.

Regarding the question of how the drive-reinforcement model can be tested at a neuronal level, synaptic inputs will have to be controlled and monitored precisely for a single neuron while the neuron's frequency of firing is continually monitored. It will be necessary to measure the direction and preferably also the magnitude of the changes in efficacy of affected synapses. Changes in synaptic inputs, as potential CSs, and changes in neuronal outputs, representing potential reinforcement, will have to be tested to determine which, if any, input and output patterns yield changes in the efficacy of synapses. In this way, it can be established whether onsets and offsets of hypothesized neuronal CSs and USs determine the efficacy of synapses in the manner specified by the drive-reinforcement model.

Experimental evidence that bears on this question of neuronal learning mechanisms has been obtained from studies involving the phenomenon of long-term potentiation (LTP). The results have been interpreted to suggest that neurons are Hebbian in character with respect to their learning mechanisms (Levy, 1985; Levy and Desmond, 1985;

Kelso, Ganong and Brown, 1986). However, the relationship of the phenomenon of LTP to learning is unclear at this time (Morris and Baker, 1984). As Bliss and Lomo (1973, p. 355) point out in the article in which they reported their discovery of LTP: "Whether or not the intact animal makes use in real life of a property which has been revealed by synchronous, repetitive volleys to a population of fibres the normal rate and pattern along which are unknown, is another matter."

Recent experimental results involving LTP suggest that sequential neuronal inputs may be more efficacious than simultaneous inputs in causing synaptic weight changes to occur. Larson and Lynch (1986) have shown that brief high frequency pulse trains delivered to nonoverlapping sets of synapses of hippocampal neurons are most effective in inducing LTP if the pulse train to a first set of synapses precedes a pulse train to a second set by 200 milliseconds. With this experimental procedure, LTP is induced only in the second set of synapses. LTP is not induced in either set of synapses if the delay is reduced to zero or extended to two seconds.

Recently, long-term depression (LTD) of parallel fiber test responses after conjunctive stimulation of parallel and climbing fiber inputs has been demonstrated in the cerebellum (Ito, Sakurai, and Tongroach, 1982; Ito, 1986). However, the relationship of this phenomenon to classical conditioning is unclear at this time because, as Byrne (1987, p. 411) notes: "Activation of parallel fiber input during the period between 20 ms prior and 150 ms after climbing fiber stimulation were roughly equivalent in inducing LTD [Ekerot and Kano,

cited in Ito, 1984]. This indicates that the neural analog of the US (climbing fiber input) can induce a change in the neural analog of the CS (parallel fiber input) even if the CS occurs after the US. Therefore the intrinsic mechanism could support backward conditioning, a phenomenon that is not observed with behavioral conditioning."

Additional experimental evidence relevant to the question of neuronal learning mechanisms has been obtained by Baranyi and Feher (1978, 1981 a, b, c) who have attempted to classically condition pyramidal neurons in the cat's motor cortex. CSs in the form of presynaptic activity were paired with USs in the form of postsynaptic cell firing. Evidence of conditioning was obtained in the form of enhanced EPSPs, with the enhancement being sustained for up to 41 minutes. The relationship of these experimental results to classical conditioning phenomena remains to be demonstrated, however, because evidence of conditioning was obtained for interstimulus intervals ranging from 0 to 400 ms and for either forward or backward pairing of the CS and US.

In summary, Baudry (1987, p. 168), in a group report from a Dahlem Workshop, offered this assessment of some of the experimental evidence discussed above: "For discrete stimulus-response learning (i.e., skeletal muscle responses), no learning occurs with backward (UCS first) or simultaneous onset or in fact until the CS precedes the UCS by nearly 100 ms. Learning is best with intervals from 200 to 400 ms and decreases as the interval is lengthened further. In terms of current models, the Aplysia system seems to follow this function remarkably well and this

seems also to be the case for Hermissenda [Lederhendler and Alkon, 1986]. It is not yet clear how LTP and LTD could satisfy this function although the newly described paradigm to obtain LTP [Larson and Lynch, 1986] also seems to follow this temporal specificity."

SECTION 6

DISCUSSION

The learning mechanism underlying nervous system function (if, indeed, there is a single basic mechanism) may not be of the character suggested by the Hebbian neuronal model. The Hebbian model suggests that approximately simultaneous neuronal signals are associated. The drive-reinforcement neuronal model, on the other hand, suggests that sequential changes in neuronal signals are associated. An implication of the drive-reinforcement model is that nervous systems, in effect, pay attention to changes, encoding causal relationships between changes as the basis for learning.

Besides psychology and neuroscience, several other disciplines have been addressing questions related to learning. These disciplines include (a) the cybernetically oriented efforts referred to as connectionist or neural network modeling, (b) artificial intelligence research, and (c) adaptive control theory and adaptive signal processing. In this section, the implications of the drive-reinforcement neuronal model for each of these approaches will be considered.

Connectionist and neural network modeling

For a few decades now, neural network models, or what are sometimes more generally referred to as connectionist models, have been proposed as theoretical models of nervous system function. Connectionist models have also been proposed as engineering solutions to problems, without any claim of biological relevance. In either case, with or without the claim

of biological relevance, the thrust of connectionist modeling has been to address the issues of memory, learning and intelligence by means of cybernetically oriented designs for massively parallel systems (Hinton and Anderson, 1981; Grossberg, 1982, 1987; Klopf, 1982; Levine, 1983; Kohonen, 1984; Barto, 1985; Feldman, 1985; Rumelhart and McClelland, 1986; McClelland and Rumelhart, 1986).

In recent years, several approaches to connectionist modeling have come to the fore, these approaches appearing to have promise in terms of solving the problem of accomplishing learning in large, deep networks. The ultimate potential of these approaches cannot be assessed yet because efforts to scale up the respective connectionist networks are only beginning. What can be done at this point and what I will attempt to do here is to assess some of the approaches for their relevance to animal learning.

One dimension along which connectionist models may be assessed has to do with the nature of the feedback the models require from their environments. Some connectionist models operate in a strictly open loop fashion, requiring no feedback from their environment. An example is the connectionist model due to Fukushima (1980, 1982). Fukushima's network, when presented with spatial patterns, adjusts connection weights so that the patterns tend to cluster in useful ways, for some purposes of pattern classification. No feedback from the environment is given or required. One question that arises is whether networks operating in this way, in an open loop or nongoal-seeking fashion, can be relevant to biological information processing. An implication of the drive-reinforcement

neuronal model and of the learning theory implied by the model is that feedback loops through the environment are a fundamental part of biological information processing. In biological systems, it appears that positive and negative feedback loops, constituting drives, support goal-seeking and that the changes in the levels of activity of these closed loops or drives constitute reinforcement.

Nearest neighbor classifications of spatial patterns, like that accomplished with Fukushima's clustering technique, can also be accomplished with Boltzmann machines (Hinton, Sejnowski, and Ackley, 1984; Ackley, Hinton, and Sejnowski, 1985; Hinton and Sejnowski, 1986) and what are sometimes called Hopfield networks (Hopfield, 1982; Cohen and Grossberg, 1983; Hopfield, 1984; Hopfield and Tank, 1985, 1986; Tesauro, 1986). These latter two classes of connectionist models, having been inspired by theoretical models in physics, utilize symmetric connections and what may be called adaptive equilibrium processes in which the networks settle into minimal energy states. The networks have been demonstrated to have interesting and potentially useful properties including, for example, in the case of Hopfield networks, solving analogs of the traveling salesman problem. However, symmetric network connections and adaptive equilibrium processes have not yet been demonstrated to be relevant to the modeling of nervous system function, at least with regard to the underlying learning mechanisms. It may be noted that a wide range of classical conditioning phenomena are predicted by the drive-reinforcement neuronal model and it uses no symmetric connections or adaptive equilibrium processes. What the

drive-reinforcement neuronal model does utilize is the real-time operation of drives and reinforcers that can be understood in terms of a network's ongoing, closed loop interactions with its environment.

Continuing to look at connectionist models in terms of the nature of the feedback they require from their environment, a class of models that might be considered the other extreme from open loop models are those using supervised learning mechanisms. Such network models require detailed feedback in the form of an error signal indicating the difference between a desired output and the network's actual output. Rosenblatt (1962), Widrow (1962), and subsequently many others have investigated connectionist models utilizing supervised learning mechanisms. For these network models, error signals play no role in a theoretical neuron's computations relative to its input-output relationship, their only role being to instruct the neuron with regard to the modification of its synaptic weights. Supervised learning mechanisms introduce the need for a "teacher" to provide a learning system with desired responses. In contrast, the drive-reinforcement neuronal model, like some other real-time learning mechanisms, does not require the introduction of a teacher and, thus, is an example of an unsupervised learning mechanism. In the case of the drive-reinforcement neuronal model, fixed (nonplastic) synapses mediating USs function like an internal teacher or goal specification.

It should be noted that unsupervised learning mechanisms have sometimes been associated with systems that operate in an open loop mode with respect to their environment. Unsupervised learning mechanisms have

also been associated with clustering techniques as an approach to pattern recognition. However, as defined here, unsupervised learning mechanisms represent that class of learning mechanisms that do not require a teacher external to the learning system and, thus, they may be utilized in learning systems that operate either in an open or closed loop mode with respect to their environment.

The distinction between unsupervised learning mechanisms that do not require a teacher and supervised learning mechanisms that do require a teacher would appear to be of fundamental importance. While supervised learning mechanisms may have a role to play in artificial intelligence, it would seem that only unsupervised learning mechanisms are likely to be relevant to the modeling of natural intelligence. In general, biological systems accomplish learning without a teacher being present in any explicit sense. Of course, a biological system's environment always functions as a teacher in an implicit sense but that is exactly what real-time unsupervised learning mechanisms can take into account, as could be seen in the results of the computer simulations of the drive-reinforcement neuronal model presented earlier.

One qualification is in order regarding the role of supervised learning mechanisms in natural intelligence. It is clear that something like supervised learning mechanisms play a large part in natural intelligence <u>at higher, cognitive levels</u>. At such levels, explicit teachers play an important role. However, I suggest that this has misled neural network modelers, causing them to introduce supervised learning mechanisms at a fundamental level. It is this hypothesized fundamental

role for supervised learning mechanisms that I think is unlikely to be valid in the case of neural network or connectionist models, if the models are intended to be relevant for natural intelligence.

Regarding connectionist models that employ supervised learning mechanisms, the most promising recent form of this class of models is due to Werbos (1974), Parker (1982, 1985), Le Cun (1985), and Rumelhart, Hinton and Williams (1985, 1986). They have proposed mechanisms for propagating error signals from the output layer back to the input layer of a network. The performance of the resulting networks has been encouraging and, therefore, the question arises of whether these connectionist models may be relevant to the understanding of animal learning. Such relevance seems unlikely for two reasons that, in part, I have already discussed. First, animals do not receive error signals during learning except, in the case of humans, after a fairly high level of cognitive function has been achieved. Second, the drive-reinforcement neuronal model demonstrates that, at least for classical conditioning phenomena that appear to be fundamental to learning, back propagating error correction mechanisms are not required.

Recognizing that animal learning does not, in general, involve evaluative feedback from the environment, some investigators have moved away from supervised learning in which error signals must be provided to the learning system. A step in the direction of unsupervised learning is reinforcement learning (Farley and Clark, 1954; Minsky, 1954; Barto and Sutton, 1981a;, 1981b; Sutton, 1984; Barto and Anandan, 1985; Barto and Anderson, 1985,) or what Widrow, Gupta, and Maitra (1973) have called

learning with a critic. Williams (1986, 1987) notes that in this type of learning the network may be provided with performance feedback as simple as a scalar signal, termed reinforcement, that indicates the network's degree of success. Reinforcement learning networks have been demonstrated to be workable (e.g., see Barto, Sutton, and Anderson, 1983), at least in the case of small scale versions. Furthermore, reinforcement learning networks appear more likely to be biologically relevant than supervised learning networks because less evaluative feedback is required from the environment. However, an implication of the drive-reinforcement neuronal model is that environmental feedback does not come in the form of reinforcement but, rather, comes in the form of changes in drive levels. Biological systems appear to compute their own reinforcement by utilizing learning mechanisms that compare current and recent drive levels. In this way, a drive-reinforcement learning mechanism requires no evaluative feedback from the environment. The environment simply provides sensory input, some of which functions as drives, and when the drive levels change, it is hypothesized that neurons and nervous systems treat these changes in drive levels as reinforcement.

Having used the expression, "evaluative feedback," I should define it. By evaluative feedback, I mean any kind of signal that requires the environment (actually, a "teacher" or "trainer" in the environment) to make some judgment about the performance of the learning system that is receiving the feedback. In an extreme case, that could mean the teacher or trainer would have to know the desired response and would then inform the learning system of the direction and magnitude of its error. In a

less extreme case, the teacher or trainer could utilize implicit or explicit criteria to form judgments about whether the learning system's performance was improving or not and then signal these evaluations of relative levels of performance to the learning system. Nonevaluative feedback, then, is any signal a learning system can generate for itself, without the aid of a teacher or trainer, simply by having an appropriate sensor with which to detect events in the environment.

Whether feedback comes to a learning system in the form of drives, reinforcers, or error signals has relevance with regard to two further questions: What should constitute the innate knowledge in a learning system and what form should the innate knowledge take? A reinforcement or a supervised learning system will, innately, know how to utilize reinforcement signals or error signals to discover appropriate drives. A drive-reinforcement learning system, on the other hand, will begin with some primary drives in place and will then acquire additional drives, utilizing changes in the current drives as reinforcers. Biological systems appear to take this latter approach, beginning with some primary or innate drives and then building acquired drives on top of them.

This approach may offer a solution to a fundamental problem in connectionist modeling. A basic question has been that of how the network elements or neurons in a large, deep, multilayered network can learn to respond properly without direct feedback from a teacher informing them of what their correct responses should have been at each step along the way. The answer suggested by drive-reinforcement learning theory, as outlined earlier, is to utilize whatever network drives

(feedback loops) are already in place and then treat changes in drive levels as reinforcers. In this way, reinforcement signals are always available locally (i.e., changes in neuronal drive levels can be computed locally) and, thus, there would appear to be no requirement for a teacher, trainer or critic at any level in the network. (This does not preclude the eventual evolution, at higher levels in a neural network, of global reinforcement centers that could aid the process of learning by providing overall direction.) Additional theoretical work including computer simulations of large, deep networks will be required to test this idea that drive-reinforcement learning mechanisms will enable multilayered networks to learn to model their environment appropriately without evaluative feedback from the environment.

Having examined the kinds of environmental feedback required by various classes of connectionist models, let us now consider the related question of what kinds of goals are implemented in these networks. In supervised learning systems, the goal is to minimize the error signal. In reinforcement learning systems, the goal may be to maximize a scalar associated with the reinforcement function. In drive-reinforcement learning systems, the goal may be to reduce drives although, as discussed in an earlier section of this report, biological systems don't always appear to be reducing drives and, even if they are, the behavioral manifestations can be subtle and complex. Some of the subtleties and complexities may be due to global reinforcement centers arising in nervous systems at the level of the limbic system and hypothalamus. Such global reinforcement centers may, in part, be responsible for certain

theorists proposing reinforcement learning systems as models of nervous system function. At a still higher level of nervous system function, cognitive processing appears to have motivated the introduction of supervised learning systems as theoretical models. From this perspective, we see that the drive-reinforcement learning mechanism might reflect the neuronal level of nervous system function, with reinforcement and supervised learning mechanisms reflecting progressively higher levels of function. It would seem then that it is important to be clear about what level of nervous system function one is modeling. Furthermore, modeling higher levels of nervous system function may require taking into account the nature of the learning mechanisms that operate at lower levels.

Regarding drive reduction as the possible goal of biological systems and, perhaps, as the goal of drive-reinforcement networks, one point that should be made is that drive reduction would seem to be the goal <u>for drives that are implemented as negative feedback loops</u>. Drives implemented as positive feedback loops would seem to support the goal of drive induction rather than drive reduction. Having said this, it may then be observed that, in the case of biological systems, drive induction, as in the pursuit of prey, always seems to be followed by drive reduction, as in the consumption of prey. This may suggest a simple general principle for the design (or evolution) of drive-reinforcement networks: primary drives implemented as positive feedback loops should always lead, when activated, to the subsequent activation of primary drives that are implemented as negative feedback

loops. If this principle is followed, then all drives will, ultimately, support the goal of drive reduction. This may help to insure the stability of learning systems.

I have traversed the conceptual or theoretical territory of connectionist models twice now, once looking at the kinds of feedback various classes of models require from their environments and once looking at the nature of the goals implemented in these models. I want to make one more pass, examining the algorithmic or heuristic character of various connectionist models.

Supervised learning mechanisms, in their most recent form, in which back propagation techniques are utilized, have a certain appeal because of what I would suggest is their nearly algorithmic character. I mean this in the mathematical sense in which an algorithm is defined to be a procedure that is guaranteed to produce a result, as distinguished from a heuristic that, like a rule of thumb, may or may not produce the desired outcome. Back propagating error correction learning mechanisms utilize gradient descent techniques such that they provide, with some allowances for the problem of getting hung up on local peaks, an optimal solution to the problem confronting the network, the problem being to arrive at the best set of connection weights. Back propagating error correction networks become of interest, then, from a theoretical standpoint, irrespective of their biological relevance, because the models may represent optimal or near optimal solutions of certain problems. Even here, there may be difficulties though, because for the larger, deeper networks many theorists are interested in, scaling up of back propagating

error correction networks may pose an obstacle (Plaut, Nowlan, and Hinton, 1986).

At any rate, if we consider that back propagating error correction networks have something of an algorithmic character, the other extreme might be connectionist networks that utilize random search techniques to identity reasonable values for the connection weights (e.g., see Barron, 1968). Random search techniques would seem to be about as far removed from an algorithmic character as a learning mechanism can get.

In between these two extremes are such classes of models as reinforcement and drive-reinforcement learning mechanisms that appear to have a heuristic character. For example, utilizing drives and reinforcers as the basis for learning may not guarantee correct results but, on the average, such an approach to learning appears to be effective in the case of biological systems.

Artificial intelligence

Fundamental to the process of learning in the case of the drive-reinforcement neuronal model is the temporal shaping of behavior. This is in contrast to the kinds of processes that occur in artificial intelligence where the emphasis is placed on what might be called cognitive searching. "Cognitive" because there is an emphasis on the rational and symbolic aspects of intelligence and "searching" because there is an emphasis on selecting from a large number of possible behaviors. An implication of the drive-reinforcement neuronal model is that, fundamentally, natural intelligence and the learning mechanisms

that support it do not involve symbols or searching but, rather, actions and shaping. Learned behavior is gradually shaped through experience to become more appropriate. This dynamic process yields associations that refine behavior that is already in place. Animals are continually "riding" a large number of feedback loops that reach through the animal and out into the environment. The more cognitively or symbolically oriented kinds of searching through large numbers of possibilities that humans sometimes engage in is, most likely, an emergent phenomenon that arises out of the internalization of a very large number of causal relations, this internalization being accomplished, it would seem, with something like a drive-reinforcement learning mechanism that temporally refines actions. Another way of saying this is that first we learn to grasp an object and then we learn to grasp a problem.

The comments I am making regarding artificial intelligence research apply as well, I feel, to cognitive science. There seems to be the view in both of these disciplines that memory, learning and intelligence have to do, fundamentally, with cognition. However, doesn't natural intelligence have to do with <u>action</u>, <u>emotion</u>, and <u>cognition</u>? The drive-reinforcement neuronal model contains what may be a complete set of the fundamental elements that underlie intelligence, namely, outputs that reflect actions, inputs and changes in inputs that reflect drives and reinforcers, synaptic weights that represent knowledge, and changes in synaptic weights that represent learning. The seeds of action, emotion, and cognition appear to be present in the drive-reinforcement neuronal model.

In such areas of artificial intelligence research as image understanding and the related area of pattern recognition (although the latter is sometimes more closely associated with connectionist models than with mainstream artificial intelligence), the tendency has been to treat the temporal aspects of intelligent information processing as too difficult for current techniques to handle. (Some recent research constitutes exceptions to this statement.) Often, ways have been sought to automatically understand static scenes or to recognize spatial patterns. The temporal aspects of natural intelligence, associated with motion and associated with real-time information processing, in general, have frequently not been addressed in image understanding and pattern recognition research, the strategy seeming to be that these difficult issues will be addressed later, when these fields of research are more advanced. But if the temporal and, indeed, real-time aspects of natural intelligence turn out to be fundamental with regard to learning, as the drive-reinforcement neuronal model suggests, could it be that the goals of image understanding and pattern recognition research will be more easily achieved if the temporal or real-time aspects of intelligent information processing are confronted first rather than last?

Having discussed cognitive searching and its role in artificial intelligence, it may be useful at this point to comment on evolutionary models of learning because such models also invoke search mechanisms in a fundamental way. Fogel, Owens, and Walsh (1966), Klopf and Gose (1969), and Holland (1975), for example, have proposed evolutionary models of learning in which alternative structures or behaviors are generated

randomly or by some process that is more systematic than a purely random one. Then, the alternatives are evaluated and the best are saved. Such an evolutionary process appears to be fundamentally different from a learning process. Fundamentally, learning does not appear to involve generating and evaluating alternatives. Rather, as discussed earlier, learning appears to involve the direct temporal shaping of behavior. Experienced causal relationships are internalized; i.e., associations are formed directly as a result of the experience. For example, when a bell rings and food follows, animals form associations directly. No search process occurs. Of course, at a higher level, searching can be occurring. It can be seen that if an animal is exploring its environment and causes a bell to ring and then food follows, the consequences of the exploratory or search process may result in the direct temporal shaping of behavior. Direct temporal shaping of behavior may be occurring then at the most fundamental level and a search process may be occurring at a higher level.

In summary, it could be said that an implication of the drive-reinforcement model is that time is the teacher (that is to say, real-time) and behavior or actions is what is taught. Ultimately, in a phylogenetically advanced organism like a human, knowledge acquisition, representation, and utilization become important too and then a process like the one I am calling cognitive searching takes on increasing importance. However, it seems that this may have misled artificial intelligence researchers and cognitive scientists, drawing their attention away from the underlying mechanisms that appear to have more to

do with temporal shaping. Artificial intelligence researchers have, for example, sometimes been dismayed by the lack of common sense in the systems they have designed. Could it be that common sense derives from the operation of drives and reinforcers and from the kind of real-time embedding in the environment that is characteristic of biological systems?

Adaptive control theory and adaptive signal processing

For several decades now, control theory has been successfully applied to the problems of analyzing and synthesizing automatic control systems. Adaptive control theory seeks to extend control system applications to cases in which adaptation or learning is required on the part of the automatic controller (e.g., see Chalam, 1987). In this way, control theory contacts the problem of learning in the context of engineering applications.

Related to the subject of adaptive control theory is adaptive signal processing (e.g., see Widrow and Stearns, 1985). In both adaptive control and adaptive signal processing, it is sometimes assumed that a "desired response" or "training signal" is available with which the controller's or signal processor's actual output can be compared for the purpose of learning. Drive-reinforcement learning theory, as outlined earlier, suggests an alternative way to extend control theory or signal processing techniques for the case of learning, such that no knowledge of a desired response or training signal is required when the learning system is operating.

In the drive-reinforcement learning theory outlined earlier, network drives are fundamental. In control theory, negative feedback loops are fundamental. But network drives, as I have defined them, and negative feedback loops are one and the same thing. (One qualification: in biological systems, network drives may also occasionally be positive feedback loops.) One sees that drive-reinforcement theory and control theory start on the same basis. It can then be seen that drive-reinforcement theory suggests a "natural" learning mechanism for control and signal processing systems. While I am not aware of any adaptive control or signal processing systems using lagged derivatives of inputs and outputs as a basis for adaptation, such a learning mechanism would seem to constitute a straightforward extension of conventional control system and signal processing techniques.

The essence of the drive-reinforcement learning mechanism, in adaptive control theoretic terms, can be simply stated. A network of drive-reinforcement neurons, viewed as a control system, will interact with its environment through some set of positive and negative feedback loops. Pursuit of prey, for example, may involve positive feedback loops, as noted earlier, and avoidance of predators may involve negative feedback loops, At any given time, a biological system will be interacting with its environment through a set of <u>actual</u> positive and negative feedback loops that constitute its current primary and acquired drives and through a set of <u>potential</u> positive and negative feedback loops that constitute possible future acquired drives. Potential acquired drives will become actual if the inputs for the potential drives

become active no more than τ time steps before any of the current actual drives change their levels of activity. In this way, what may be called a drive-reinforcement controller will learn to control its output not only to deliver more or less of a control signal (as current adaptive controllers do) but also to deliver the control signal sooner or later. That is to say, a drive-reinforcement controller would be expected to modify not only the amplitudes of its responses but also the timing.

Memory and learning

Before concluding this discussion of some of the implications of the drive-reinforcement neuronal model, a few words should be said about memory and how it relates to learning. As Squire (1986) notes, in phylogenetically old animals such as invertebrates, what is learned takes the form of procedural memories. In phylogenetically recent animals such as mammals, what is learned can also take the form of declarative memories. The distinction between procedural and declarative memories is that between skills and procedures, on the one hand, and specific facts and data, on the other.

The drive-reinforcement learning mechanism appears to be well suited for the laying down of procedural memories because the learning mechanism treats time as a fundamental dimension, utilizing time derivatives of the neuronal inputs and outputs and correlating the derivatives across a temporal interval. If the drive-reinforcement learning mechanism should turn out to be the learning mechanism for acquiring procedural memories, could it also turn out to be the learning mechanism for acquiring declarative memories? To see how this could be a possibility, it may be

necessary to consider the interaction of the brain's attention mechanism with the registration of sensory and other information in the cerebral cortex. The medial temporal cortex and especially the hippocampal formation and associated areas appear to be important with respect to declarative memories. Squire (1986) notes that the capacity for declarative memories reaches its greatest development in mammals in which these cortical structures are most fully elaborated. Given our tendency to remember that to which we attend, might it be that signals generated by the attention mechanism, the signals originating perhaps in the thalamic reticular formation (Klopf, 1982), interact with sensory and other information registering in the medial temporal cortex, such that the temporal relationships specified by the drive-reinforcement learning mechanism are satisfied and declarative memories result? In general, could the role of the attention mechanism in the laying down of both procedural and declarative memories be the induction of Δy's at appropriate times relative to Δx's so that the resulting synaptic weight changes yield learning?

SECTION 7

CONCLUDING REMARKS

In the Foreword to Olds' (1977) book on Drives and Reinforcements, Neal Miller remarks (p. v): "A fundamental step in the line of evolution leading to human behavior was the development of learning, a new process of adaptation that could occur far more rapidly within the lifetime of the individual instead of slowly during the evolution of the species. In determining which particular response will be performed and learned, the selective factor is reinforcement which, in turn, is closely related to the drives that are active at a given time." In this report, I have attempted to relate drives and reinforcers by means of a theoretical model of neuronal function. The model has been demonstrated to predict a wide range of classical conditioning phenomena. Implications of the model have been considered for the fields of animal learning theory, connectionist and neural network modeling, artificial intelligence research, adaptive control theory, and adaptive signal processing. It has been concluded that a real-time learning mechanism that does not require evaluative feedback from the environment may be fundamental to natural intelligence and that such a learning mechanism may have implications for artificial intelligence.

In addition to accomplishing experimental tests of the neuronal model, a useful next step may be to simulate networks of the proposed theoretical neurons to determine the properties of the networks, in general, and, in particular, to determine if instrumental conditioning

phenomena emerge. Actually, in pursuing this theoretical work, it may be useful to simulate not only the neural network but also a simplified organism controlled by the neural network and a simplified environment with which the organism is interacting. [See Barto and Sutton (1981b) for an example of how this kind of simulation can be carried out.] By means of such computer simulations of nervous systems, organisms, and environments, it may become possible to make behavioral observations on a mathematically well defined network of drive-reinforcement neurons during the process of learning.

REFERENCES

Ackley, D. H., Hinton, G. H., & Sejnowski, T. J. (1985). A learning algorithm for Boltzman machines. Cognitive Science, 9, 147-169.

Anderson, J. A., Silverman, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. Psychological Review, 84, 413-451.

Ashton, A. B., Bitgood, S. C., & Moore, J. W. (1969). Differential conditioning of the rabbit nictitating membrane response as a function of US shock intensity and duration. Psychonomic Science, 15, 127-218.

Ayres, J. J. B., Albert, M., & Bombase, J. C. (1967). Extending conditioned stimuli before versus after unconditioned stimuli: Implications for real-time models of conditioning. Journal of Experimental Psychology: Animal Behavior Processes, 13(2), 168-181.

Baker, T. W. (1968). Component strength in a compound CS as a function of number of acquisition trials. Journal of Experimental Psychology, 79, 347-352.

Balsam, P. D., & Schwartz, A. L. (1981). Rapid contextual conditioning in autoshaping. Journal of Experimental Psychology: Animal Behavior Processes, 7, 382-393.

Baranyi, A., & Ferher, O. (1978). Conditioned changes of synaptic transmission in the motor cortex of the cat. Experimental Brain Research, 33, 283-298.

Baranyi, A., & Ferher, O. (1981a). Long-term facilitation of excitatory synaptic transmission in single motor cortical neurones of the cat produced by repetitive pairing of synaptic potentials and action potentials following intracellular stimulation. Neuroscience Letters, 23, 303-308.

Baranyi, A., & Ferher, O. (1981b). Synaptic facilitation requires paired activation of convergent pathways in the neocortex. Nature (London), 290, 413-415.

Baranyi, A., & Ferher, O. (1981c). Intracellular studies on cortical synaptic plasticity: conditioning effect of antidromic activation on test-EPSPs. Experimental Brain Research, 41, 124-134.

Barron, R. L. (1968). Self-organizing and learning control systems. In H. L. Ostreicher & D. Moore (Eds.), Cybernetic Problems in Bionics (pp. 147-203). New York: Gordon and Breach.

Barto, A. G. (1985). Learning by statistical cooperation of self-interested neuron-like computing elements. Human Neurobiology, 4, 229-256.

Barto, A. G., & Anandan, P. (1985). Pattern recognizing stochastic learning automata. IEEE Transactions on Systems, Man and Cybernetics, 15, 360-374.

Barto, A. G., & Anderson, C. W. (1985). Structural learning in connectionist systems. Proceedings of the Seventh Annual Conference of the Cognitive Science Society, Irvine, California, 43-53.

Barto, A. G., & Sutton, R. S. (1981a). Goal seeking components for adaptive intelligence: An initial assessment (Technical Report 81-1070). Wright-Patterson Air Force Base, Ohio: Air Force Wright Aeronautical Laboratories (DTIC Report AD A101476 available from the Defense Technical Information Center, Cameron Station, Alexandria, Virginia 22304-6145).

Barto, A. G., & Sutton, R. S. (1981b). Landmark learning: An illustration of associative search. Biological Cybernetics, 42, 1-8.

Barto, A. G., & Sutton, R. S. (1982). Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element. Behavioural Brain Research, 4, 221-235.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike elements that can solve difficult learning control problems. IEEE Transactions on Systems, Man, and Cybernetics, 13, 835-846.

Baudry, M. (1987). Activity-dependent regulation of synaptic transmission and its relationship to learning (Group Report, Dahlem Workshop). In J.-P. Changeux & M. Konishi (Eds.), The Neural and Molecular Bases of Learning (pp. 152-175). New York: Wiley.

Berlyne, D. E. (1973). The vicissitudes of aplopathematic and thelematoscopic pneumatology (or the hydrography of hedonism). In D. E. Berlyne & K. B. Madsen (Eds.), Pleasure, Reward, Preference: Their Nature, Determinants, and Role in Behavior (pp. 1-33). New York: Academic Press.

Bindra, D. (1968). Neuropsychological interpretation of the effects of drive and incentive-motivation on general activity and instrumental behavior. Psychological Review, 75, 1-22.

Bindra, D. (1976). A theory of intelligent behavior. New York: Wiley.

Bindra, D. (1967). How adaptive behavior is produced: a perceptual-motivational alternative to response-reinforcement. The Behavioral and Brain Sciences, 1, 41-91.

Blazis, D. E. J., Desmond, J. E., Moore, J. W., & Berthier, N. E. (1986). Simulation of the classically conditioned nictitating membrane response by a neuron-like adaptive element: A real-time variant of the Sutton-Barto model. Proceedings of the Eighth Annual Conference of the Cognitive Science Society, Amherst, Massachusetts, 176-186.

Blazis, D. E. J., & Moore, J. W. (1978). Simulation of a classically conditioned response: Components of the input trace of a cerebellar neural network implementation of the Sutton-Barto-Desmond model (Coins Technical Report 87-74). Amherst, Massachusetts: University of Massachusetts.

Bliss, T. V. P., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetised rabbit following stimulation of the perforant path. Journal of Physiology (London), 232, 331-356.

Bolles, R. C. (1967). Theory of motivation. New York: Harper and Row.

Bolles, R. C. (1972). Reinforcement, expectancy and learning. Psychological Review, 79, 394-409.

Booker, L. B. (1982). Intelligent behavior as an adaptation to the task environment. (Doctoral dissertation, University of Michigan, 1982). Dissertation Abstracts International, 43, 4696.

Brown, P. L., & Jenkins, H. M. (1968). Autoshaping of the pigeon's key peck. Journal of the Experimental Analysis of Behavior, 11, 1-8.

Burkhardt, P. E. & Ayres, J. J. B. (1978). CS and US duration effects in one-trial simultaneous fear conditioning as assessed by conditioned suppression of licking in rats. Animal Learning and Behavior, 6, 225-230.

Byrne, J. H. (1987). Cellular analysis of associative learning. Physiological Reviews, 67(2), 329-439.

Calvin, W. H. (1975). Generation of spike trains in CNS neurons. Brain Research, 84, 1-22.

Carew, T. J., Hawkins, R. D., Abrams, T. W., & Kandel, E. R. (1984). A test of Hebb's postulate at identified synapses which mediate classical conditioning in Aplysia. Journal of Neuroscience, 4, 1217-1224.

Carew, T. J., Hawkins, R. D., & Kandel, E. R. (1983). Differential classical conditioning of a defensive withdrawal reflex in Aplysia Californica. Science, 219, 397-400.

Carew, T. J., & Sahley, C. L. (1986). Invertebrate learning and memory: From behavior to molecules. Annual Review of Neuroscience, 9, 435-487.

Carew, T. J., Walters, E. T. & Kandel, E. R. (1981). Classical conditioning in a simple withdrawal reflex in Aplysia Californica. Journal of Neuroscience, 1, 1426-1437.

Chalam, V. V. (1987). Adaptive control systems. New York: Marcel Dekker.

Cohen, M. A., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transactions on Systems, Man and Cybernetics, SMC-13, No. 5, 815-826. Courillon, P. A., & Litterman, M. E. (1982). Compound conditioning in honeybees. Journal of Comparative Psychology, 96, 192-199.

Delgado, J. M. R., Roberts, W. W., & Miller, N. E. (1954). Learning motivated by electrical stimulation of the brain. American Journal of Physiology, 179, 587-593.

Dickinson, A., & Mackintosh, N. J. (1978). Classical conditioning in animals. Annual Review of Psychology, 29, 587-612.

Dolan, J. C., Shishimi, A., & Wagner, A. R. (1985). The effects of signaling the US in backward conditioning: A shift from excitatory to inhibitory learning. Animal Learning and Behavior, 13(3), 209-214.

Donegan, N. H., & Wagner, A. R. (1987). Conditioned diminution and facilitation of the UR: A sometimes opponent-process interpretation. In I. Gormezano, W. F. Prokasy, & R. F. Thompson (Eds.), Classical Conditioning (pp. 339-369). Hillsdale: Erlbaum.

Eccles, J. C. (1964). The physiology of synapses. New York: Academic Press.

Estes, W. K. (1950). Toward a statistical theory of learning. Psychological Review, 57, 94-107.

Farley, B. G., & Clark, W. A. (1954). Simulation of self-organizing systems by digital computer. IRE Transactions on Information Theory, PGIT-4, 76-84.Farley, J., & Alkon, D. L. (1985). Cellular mechanisms of learning, memory, and information storage. Annual Review of Psychology, 36, 419-494.

Feldman, J. A. (Ed.) (1985). Special issue on connectionist models and their applications. Cognitive Science, 9.

Flaherty, C. F. (1985). Animal learning and cognition. New York: Knopf.

Flaherty, C. F., Hamilton, L. W., Gandelman, R. J., & Spear, N. E. (1977). Learning and memory. Chicago: Rand McNally College Publishing Company.

Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). Artificial intelligence through simulated evolution. New York: Wiley.

Fowler, H. (1971). Implications of sensory reinforcement. In R. Glaser (Ed.), The Nature of Reinforcement (pp. 151-195). New York: Academic Press.

Freud, S. (1895, 1964). Untitled paper. In J. Strachey (Ed.), The Standard Edition of the Complete Psychological Works of Sigmund Freud (Vol. 1, pp. 281-387). New York: MacMillan.

Frey, P. W., & Ross, L. E. (1968). Classical conditioning of the rabbit eyelid response as a function of interstimulus interval. Journal of Comparative and Physiological Psychology, 65, 246-250.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36, 193-202.

Fukushima, K., & Miyake, S. (1982). Neocognition: A self-organizing neural network model for a mechanism of visual pattern recognition. In S. Amari & M. A. Arbib (Eds.), Competition and Cooperation in Neural Nets (pp. 267-285). New York: Springer-Verlag.

Fuxe, K. (1965). Acta Physiologica Scandinavica, 64 (Suppl. 247), 37-84.

Gelperin, A., Hopfield, J. J., & Tank, D. W. (1985). The logic of Limax learning. In A. Selverston (Ed.), Model Neural Networks and Behavior (pp. 237-261). New York: Plenum Press.

Gluck, M. A., & Thompson, R. F. (1987). Modeling the neural substrates of associative learning and memory: A computational approach. Psychological Review, 94 (2), 176-191.

Gormezano, I. (1972). Investigations of defense and reward conditioning in the rabbit. In A. H. Black & W. F. Prokasy (Eds.), Classical Conditioning II: Current Research and Theory (pp. 151-181). New York: Appleton.

Gormezano, I., Kehoe, E. J., & Marshall, B. S. (1983). Twenty years of classical conditioning research with the rabbit. In J. M. Sprague & A. N. Epstein (Eds.) Progress in Psychobiology and Physiological Psychology (pp. 198-273), New York: Academic Press.

Gould, J. L. (1986). The biology of learning. Annual Review of Psychology, 37: 163-192.

Gray, J. A. (1975). Elements of a two-process theory of learning. New York: Academic Press.

Grossberg, S. (1982). Studies of mind and brain. Boston: Reidel.

Grossberg, S. (Ed.) (1987). The adaptive brain (Volumes I and II). New York: North-Holland.

Guthrie, E. R. (1935). The psychology of learning. New York: Harper and Row.

Hampson, S. E. (1984). A neural model of adaptive behavior. (Doctoral dissertation, University of California, Irvine, 1983). Dissertation Abstracts International, 44, 3457B.

Hawkins, R. D., Abrams, T. W., Carew, T. J., & Kandel, E. R. (1983). A cellular mechanism of classical conditioning in Aplysia: Activity-dependent amplification of presynaptic facilitation. Science, 219, 400-405.

Hawkins, R. D., & Kandel, E. R. (1984). Steps toward a cell-biological alphabet for elementary forms of learning. In G. Lynch, J. L. McGaugh, & N. M. Weinberger (Eds.), Neurobiology of Learning and Memory (pp. 385-404). New York: Guilford Press.

Hawkins, R. D., Carew, T. J., & Kandel, E. R. (1986). Effects of interstimulus interval and contingency on classical conditioning of the Aplysia siphon withdrawal reflex. Journal of Neuroscience, 6, 1695-1701.

Hebb, D. O. (1949). The organization of behavior. New York: Wiley.

Helson, H. (1964). Adaptation-level theory. New York: Harper & Row.

Heth, C. D. (1976). Simultaneous and backward fear conditioning as a function of number of CS-UCS pairings. Journal of Experimental Psychology: Animal Behavior Processes, 2 (2), 117-129.

Hilgard, E. R., & Bower, G. H. (1975). Theories of learning (fourth edition). Englewood Cliffs: Prentice-Hall.

Hinton, G. E., & Anderson, J. A. (Eds.) (1981). Parallel models of associative memory, Hillsdale: Erlbaum.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In E. E. Rumelhart & J. L. McClelland (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations (pp. 282-317). Cambridge: Bradford Books/MIT Press.

Hinton, G. E., Sejnowski, T. J., & Ackley, D. H. (1984). Boltzmann machines: Constraint satisfaction networks that learn (Technical Report CMU-CS-84-119). Pittsburgh, Pennsylvania: Carnegie-Mellon University.

Holland, J. H. (1975). Adaptation in natural and artificial systems. Ann Arbor: University of Michigan Press.

Holland, J. H. (1986). Escaping brittleness: The possibilities of general-purpose learning algorithms applied to parallel rule-based systems. In R. S. Michalski, J. G. Garbonell, & T. M. Mitchell (Eds), Machine Learning: An Artificial Intelligence Approach, Volume II (pp. 593-623). Los Altos: Morgan-Kaufmann.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences USA, 79, 2554-2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Sciences USA, 81, 3088-3092.

Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. Biological Cybernetics, 52, 141-152.

Hopfield, J. J., & Tank, D. W. (1986). Computing with neural circuits: A model. Science, 233, 625-633.

Hull, C. L. (1943). Principles of behavior. New York: Appleton-Century-Crofts.

Hull, C. L. (1952). A behavior system: An introduction to behavior theory concerning the individual organism. New Haven: Yale University Press.

Ito, M. (1984). The modifiable neuronal network of the cerebellum. Japanese Journal of Physiology, 34, 781-792.

Ito, M. (1986). Long-term depression as a memory process in the cerebellum. Neuroscience Research, 3, 531-539.

Ito, M., Sakurai, M., & Tongroach, P. (1982). Climbing fibre-induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells. Journal of Physiology (London), 324, 113-134.

Jenkins, H. M., & Moore, B. R. (1973). The form of the autoshaped response with food or water reinforcers. Journal of the Experimental Analysis of Behavior, 20, 163-181.

Kamin, L. J. (1965). Temporal and intensity characteristics of the conditioned stimulus. In W. F. Prokasy (Ed.), Classical Conditioning: A Symposium (pp. 118-147). New York: Appleton-Century- Crofts.

Kamin, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), Miami Symposium on the Prediction of Behavior: Aversive Stimulation (pp. 9-31). Miami: University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), Punishment and Aversive Behavior (pp. 279-296). New York: Appleton-Century- Crofts.

Kandel, E. R. & Spencer, W. A. (1968). Cellular neurophysiological approaches in the study of learning. Physiological Reviews, 48 (1), 65-134.

Keene, J. J. (1975). Reward-associated excitation and pain-associated inhibition lasting seconds in rat medial pallidal units. Experimental Neurology, 49, 97-114.

Kelso, S. R. & Brown, T. H. (1986). Differential conditioning of associative synaptic enhancement in hippocampal brain slices. Science, 232, 85-87.

Kelso, S., Ganong, A. & Brown, T. H. (1986). Hebbian synapses in hippocampus. Proceedings of the National Academy of Sciences USA, 83, 5326-5330.

Klopf, A. H. (1972). Brain function and adaptive systems - a heterostatic theory [Report No. 133 (AFCRL-72-0164)]. L. G. Hanscom Field, Bedford, Massachusetts: Air Force Cambridge Research Laboratories (DTIC Report AD 742259 available from the Defense Technical Information Center, Cameron Station, Alexandria, Virginia 22304-6145).

Klopf, A. H. (1975). A comparison of natural and artificial intelligence. Association for Computing Machinery Special Interest Group on Artificial Intelligence Newsletter, No. 52.

Klopf, A. H. (1979). Goal-seeking systems from goal-seeking components: Implications for AI. Cognition and Brain Theory Newsletter, 3 (2), 54-62.

Klopf, A. H. (1982). The hedonistic neuron: A theory of memory, learning, and intelligence. Washington: Hemisphere/Harper & Row.

Klopf, A. H. (1986). A drive-reinforcement model of single neuron function: An alternative to the Hebbian neuronal model. In J. S. Denker (Ed.), Neural Networks for Computing. AIP Conference Proceedings 151, New York: American Institute of Physics, 265-270.

Klopf, A. H., & Gose, E. (1969). An evolutionary pattern recognition network. IEEE Transactions on Systems Science and Cybernetics, SSC-5 (3), 247-250.

Kohonen, T. (1977). Associative memory: A system theoretic approach. New York: Springer.

Kohonen, T. (1984). Self-organization and associative memory. New York: Springer.

Kosko, B. (1986). Differential Hebbian learning. In J. S. Denker (Ed.), Neural Networks for Computing. AIP Conference Proceedings 151, New York: American Institute of Physics, 277-282.

Kremer, E. F. (1978). The Rescorla-Wagner model: Losses in associative strength in compound conditioned stimuli. Journal of Experimental Psychology: Animal Behavior Processes, 4, 22-36.

Larson, J., & Lynch, G. (1986). Induction of synaptic potentiation in hippocampus by patterned stimulation involves two events. Science, 232, 985-988.

Le Cun, Y. (1985). One procedure d'apprentissage pour reseau a seuil assymetrique [A learning procedure for an assymetric threshold network]. Proceedings of Cognitiva, 85 (pp. 559-604), Paris, France.

Lederhendler, I., & Alkon, D. L. (1986). Temporal specificity of the CS-UCS interval for Hermissenda Pavlovian conditioning. Society for Neuroscience Abstracts, Vol. 12, Part 1, p. 40.

Levine, D. S. (1983). Neural population modeling and psychology: A review. Mathematical Biosciences, 66, 1-86.

Levy, W. B. (1985). Associative changes at the synapse: LTP in the hippocampus. In W. B. Levy, J. A. Anderson, & S. Lehmkuhle (Eds.), Synaptic modification, neuron selectivity, and nervous system organization (pp. 5-33). Hillsdale: Erlbaum.

Levy, W. B., & Desmond, N. L. (1985). The rules of elemental synaptic plasticity. In W. B. Levy, J. A. Anderson & L. Lehmkuhle (Eds.), Synaptic modification, neuron selectivity, and nervous system organization (pp. 105-121). Hillsdale: Erlbaum.

Lynch, J. C., Mountcastle, V. B., Talbot, W. H., & Yin, T. C. T. (1977). Parietal lobe mechanisms for directed visual attention. Journal of Neurophysiology, 40, 362-389.

Mackintosh, N. J. (1974). The psychology of animal learning. New York: Academic Press.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. Psychological Review, 82, 276-298.

Mahoney, W. J., & Ayres, J. J. B. (1976). One trial simultaneous and backward fear conditioning as reflected in conditioned suppression of licking in rats. Animal Learning and Behavior, 4, 357-362.

McAllister, W. R. (1953). Eyelid conditioning as a function of the CS-US interval. Journal of Experimental Psychology, 45, 417-422.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). The achievement motive. New York: Appleton.

McClelland, J. L., & Rumelhart, D. E. (Eds.) (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models. Cambridge: Bradford Books/MIT Press.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5, 115-137. Reprinted in McCulloch, W. S. (1965). Embodiments of Mind (pp. 19-39). Cambridge: MIT Press.

McNaughton, B. L., Barnes, C. A., & Rao, G. (1984). Presynaptic versus postsynaptic control over long-term synaptic enhancement. In G. Lynch, J. L. McGaugh & N. M. Weinberger (Eds.), Neurobiology of Learning and Memory (pp. 466-469). New York: Guilford Press.

Miller, N. E. (1951). Learnable drives and rewards. In S. S. Stevens (Ed.), Handbook of Experimental Psychology (pp. 435-472). New York: Wiley.

Miller, R. R., & Balaz, M. A. (1981). Differences in adaptiveness between classically conditioned responses and instrumentally conditioned responses. In N. E. Spear & R. R. Miller (Eds.), Information Processing in Animals: Memory Mechanisms (pp. 49-80). Hillsdale: Erlbaum.

Minsky, M. L. (1954). Theory of neural-analog reinforcement systems and its application to the brain-model problem. Unpublished doctoral dissertation, Princeton University, Princeton.

Mis, F. W., & Moore, J. W. (1973). Effect of preacquisition UCS exposure on classical conditioning of the rabbit's nictitating membrane response. Learning and Motivation, 4, 108-114.

Mishkin, M., & Petri, H. L. (1984). Memories and habits: Some implications for the analysis of learning and retention. In L. R. Squire & N. Butters (Eds.), Neuropsychology of Memory (pp. 287-296). New York: Guilford Press.

Mogenson, G. J., & Phillips, A. G. (1976). Motivation: A physiological construct in search of a physiological substrate. In J. M. Sprague & A. N. Epstein (Eds.), Progress in Psychobiology and Physiological Psychology (pp. 189-243). New York: Academic Press.

Moore, J. W. (1979). Brain processes and conditioning. In A. Dickinson & R. A. Boakes (Eds.), Mechanisms of Learning and Motivation: A Memorial Volume to Jerzy Konorski (pp. 111-142). Hillsdale: Erlbaum.

Moore, J. W., Desmond, J. E., Berthier, N. E., Blazis, D. E. J., Sutton, R. S., & Barto, A. G. (1986). Simulation of the classically conditioned nictitating membrane response by a neuron-like adaptive element: Response topography, neuronal firing, and interstimulus intervals. Behavioural brain Research, in press.

Moore, J. W., & Gormezano, I. (1977). Classical conditioning. In M. H. Marx & M. E. Bunch (Eds.), Fundamentals and Applications of Learning (pp. 87-120). New York: Macmillan.

Moore, J. W., & Stickney, K. J. (1980). Formation of attentional-associative networks in real time: Role of the hippocampus and implications for conditioning. Physiological Psychology, 8 (2), 207-217.

Moore, J. W., & Stickney, K. J. (1985). Antiassociations: Conditioned inhibition in attentional-associative networks. In R. R. Miller & N. E. Spear (Eds.), Information Processing in Animals: Conditioned Inhibition (pp. 209-232). Hillsdale: Erlbaum.

Morris, R., & Baker, M. (1984). Does long-term potentiation/synaptic enhancement have anything to do with learning or memory? In L. R. Squire & N. Butters (Eds.), Neuropsychology of Memory (pp. 521-535). New York: Guilford Press.

Mowrer, O. H. (1960). Learning theory and behavior. New York: Wiley (Krieger Edition, 1973).

Mpitsos, G. J., Collins, S. D., & McClelland, A. D. (1978). Learning: A model system for physiological studies. Science, 199, 497-506.

Olds, J. (1977). Drives and reinforcements: behavioral studies of hypothalamic functions. New York: Raven Press.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. Journal of Comparative and Physiological Psychology, 47, 419-427.

Parker, D. B. (1982). Learning Logic. (Invention Report S81-64, File 1). Stanford, California: Office of Technology Licensing, Stanford University.

Parker, D. B. (1985). Learning-logic (Technical Report-47). Cambridge, Massachusetts: Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology.

Pavlov, I. P. (1927). Conditioned reflexes. Oxford: Oxford University Press (Dover Edition, 1960).

Pavlov, I. P. (1928). Lectures on conditioned reflexes, New York: International Publishers.

Plaut, D. C., Nowlan, S. J., & Hinton, G. E. (1986). Experiments on learning by back propagation (Technical Report CMU-CS-86-126). Pittsburgh, Pennsylvania: Department of Computer Science, Carnegie-Mellon University.

Premack, D. (1959). Toward empirical behavior laws: I. Positive reinforcement. Psychological Review, 66, 219-233.

Rashevsky, N. (1938). Mathematical biophysics. Chicago: University of Chicago Press.

Reiss, S., & Wagner, A. R. (1972). CS habituation produces a "latent inhibition effect" but no active conditioned inhibition. Learning and Motivation, 3, 237-245.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. Psychological Bulletin, 72 (2), 77-94.

Rescorla, R. A. (1971). Summation and retardation tests of latent inhibition. Journal of Comparative and Physiological Psychology, 75, 77-81.

Rescorla, R. A. (1971). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. Learning and Motivation, 2, 113-123.

Rescorla, R. A. (1980). Pavlovian second-order conditioning: Studies in associative learning. New York: Erlbaum/Wiley.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical

Conditioning II: Current Research and Theory (pp. 64-99). New York: Appleton-Century-Crofts.

Roberts, E. (1980). Epilepsy and antiepileptic drugs: a speculative synthesis. In G. H. Glaser, J. K. Penry & D. M. Woodbury (Eds.), Antiepileptic Drugs: Mechanisms of Action (pp. 667-713). New York: Raven Press.

Rolls, E. T. (1987). Information representation, processing, and storage in the brain: Analysis at the single neuron level. In J.-P. Changeux & M. Konishi (Eds.), The Neural and Molecular Bases of Learning (pp. 503-540). New York: Wiley.

Rosenblatt, F. (1962). Principles of neurodynamics. New York: Spartan Books.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). Learning internal representations by error propagation (ICS Report 8506). San Diego, California: Institute for Cognitive Science, University of California.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations (pp. 318-364). Cambridge: Bradford Books/MIT Press.

Rumelhart, D. E., & McClelland, J. L. (Eds.) (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations (pp. 318-364). Cambridge: Bradford Books/MIT Press.

Russell, I. S. (1966). Animal learning and memory. In D. Richter (Ed.), Aspects of Learning and Memory (pp. 121-171). New York: Basic Books.

Schmajuk, N. A., & Moore, J. W. (1985). Real-time attentional models for classical conditioning and the hippocampus. Physiological Psychology, 13 (4), 278-290.

Schwartz, B. (1978, 1984). Psychology of learning and behavior. New York: Norton.

Smith, M. C., Coleman, S. R., & Gormezano, I. (1969). Classical conditioning of the rabbit's nictitating membrane response. Journal of Comparative and Physiological Psychology, 69, 226-231.

Solomon, P. R., Brennan, G., & Moore, J. W. (1974). Latent inhibition of the rabbit's nictitating membrane response as a function of CS intensity. Bulletin of the Psychonomic Society, 4, 445-448.

Solomon, R. L. (1981). Prologue. In N. E. Spear & R. R. Miller (Eds.), Information Processing in Animals: Memory Mechanisms. Hillsdale: Erlbaum.

Spence, K. W. (1956). Behavior theory and conditioning. New Haven: Yale University Press.

Squire, L. R. (1986). Mechanisms of memory. Science, 232, 1612-1619.

Stein, L., & Belluzzi, J. D. (in press). Operant conditioning of individual neurons. In M. Commons, R. Church, J. Stellar, & A. Wagner (Eds.), Quantitative Analyses of Behavior, Volume 7: Biological Determinants of Reinforcement and Memory. Hillsdale: Erlbaum.

Stein, L., Wise, C. D., & Belluzzi, J. D. (1977). Neuropharmacology of reward and punishment. In L. L. Iverson, S. D. Iversen & S. H. Snyder

(Eds.), Handbook of Psychopharmacology (pp. 25-53). Volume 8, New York: Plenum.

Sutherland, N. S., & Mackintosh, M. J. (1971). Mechanisms of animal discrimination learning. New York: Academic Press.

Sutton, R. S. (1984). Temporal credit assignment in reinforcement learning. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Sutton, R. S. (1987). Learning to predict by the methods of temporal differences (Technical Report 87-509.1). Waltham, Massachusetts: GTE Laboratories.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. Psychological Review, 88, 135-170.

Sutton, R. S., & Barto, A. G. (1987). A temporal-difference model of classical conditioning (Technical Report 87-509.2). Waltham, Massachusetts: GTE Laboratories.

Tesauro, G. (1986). Simple neural models of classical conditioning. Biological Cybernetics, 55, 187-200.

Thompson, R. F. (1976). The search for the engram. American Psychologist, 31, 209-227.

Thompson, R. F., Berger, T. W., & Madden, IV, J. (1983). Cellular processes of learning and memory in the mammalian CNS. Annual Review of Neuroscience, 6, 447-491.

Thompson, R. F., McCormick, D. A., Lavond, D. G., Clark, G. A., Kettner, R. E., & Mauk, M. D. (1983). The engram found? Initial localization of the memory trace for a basic form of associative learning. In J. M. Sprague & A. N. Epstein (Eds.), Progress in Psychobiology and Physiological Psychology (pp. 167-196). New York: Academic Press.

Thorndike, E. L. (1911). Animal intelligence. New York: MacMillan.

Toates, F. (1985). Psychobiology. (Review of The Neurobiology of Motivation and Reward by J. R. Stellar & E. Stellar). Science, 229, 962-963.

Tolman, E. C. (1932). Purposive behavior in animals and men. New York: Appleton-Century-Crofts. Reprinted, University of California Press, 1949.

Wagner, A. R. (1971). Elementary associations. In H. H. Kendler & J. T. Spence (Eds.), Essays in Neobehaviorism: A Memorial Volume to Kenneth W. Spence (pp. 187-213). New York: Appleton-Century-Crofts.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), Information Processing in Animals: Memory Mechanisms (pp. 5-47). Hillsdale: Erlbaum.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. Journal of Experimental Psychology, 76, 171-180.

Wagner, A. R., & Terry, W. S. (1975). Backward conditioning to a CS following an expected vs. a surprising UCS. Animal Learning & Behavior, 3 (4), 370-374.

Walters, E. T., & Byrne, J. H. (1983). Associative conditioning of single neurons suggests a cellular mechanism for learning. Science, 219, 405-408.

Werbos, P. J. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. Unpublished doctoral dissertation, Harvard University, Cambridge.

Widrow, B. (1962). Generalization and information storage in networks of adaline neurons. In M. C. Yovits, G. T. Jacobi & G. D. Goldstein (Eds.), Self-Organizing Systems-1962 (pp. 435-461). Washington: Spartan Books.

Widrow, B., Gupta, N. K., & Maitra, S. (1973). Punish/reward: Learning with a critic in adaptive threshold systems. IEEE Transactions on Systems, Man, and Cybernetics, 5, 455-465.

Widrow, B., & Stearns, S. D. (1985). Adaptive signal processing. Englewood Cliffs: Prentice-hall.

Williams, R. J. (1986). Reinforcement learning in connectionist networks: A mathematical analysis (Technical Report 8605). University of California, San Diego: Institute for Cognitive Science.

Williams, R. J. (1987). Reinforcement-learning connectionist systems (Technical Report NU-CCS-87-3). Boston, Massachusetts: Northeastern University.

Witten, I. H. (1977). An adaptive optimal controller for discrete-time markov environments. Information and Control, 34, 286-295.

Woody, C. D. (1982). Memory, learning and higher function. New York: Springer-Verlag.

Woody, C. D. (1986). Understanding the cellular basis of memory and learning. Annual Review of Psychology, 37, 433-493.

Zipser, D. (1986). A model of hippocampal learning during classical conditioning. Behavioral Neuroscience, 100 (5), 764-776.

APPENDIX: Parameter Specifications for the Computer Simulations of the Neuronal Models

Drive-reinforcement model

Learning rate constants: c_1=5.0, c_2=3.0, c_3=1.5, c_4=0.75, c_5=0.25(τ=5)

CS initial synaptic weight values [i.e., $w_i(t)$ at t=0]: +0.1 (excitatory weights), -0.1 (inhibitory weights). Exceptions: For the simulations reported in Figures 12 and 18, the initial values of the inhibitory synaptic weights were 0.0, thus preventing the inhibitory weights from changing during these simulations. This was done to simplify the graphs and to focus attention on the excitatory weights that were primarily responsible for the phenomena being manifested. Had the initial inhibitory synaptic weight values for Figures 12 and 18 been set at -0.1, as was done for the other simulations, small changes in inhibitory weights would have been observed at some points in these simulations while the overall phenomena being manifested would have remained unchanged.

US (nonplastic) synaptic weight values: +1.0 (excitatory weight) and 0.0 (inhibitory weight).

Lower bound on synaptic weights: $|w_i(t)| \geq 0.1$

Neuronal output limits: $0.0 \leq y(t) \leq 1.0$

Neuronal threshold: θ =0.0

CS amplitudes (measured relative to zero-level baseline): 0.2 except for Figure 7 where the amplitudes were 1.0, 0.5, and 0.25 for CS_1, CS_2, and CS_3, respectively, and Figure 17 where the amplitudes were 0.2, 0.2, and 0.4 for CS_1, CS_2, and CS_3, respectively.

US amplitudes (measured relative to zero-level baseline): 0.5 except for Figure 8 where the US amplitudes were 1.0, 0.5, and 0.25 for the USs occurring in conjunction with CS_1, CS_2, and CS_3, respectively.

CS and US timing: See Table 1 for times of onset and offset of CSs and USs within a trial. Also specified in Table 1 are the trials during which each CS and US was present. For all of the CS-US configurations, the time of onset of the first stimulus was arbitrarily chosen to be 10. Onset of a stimulus at time step, t, means that the stimulus was on during time step, t, and was not on during the previous time step. Offset of a stimulus at time step, t, means that the stimulus was off during time step, t, and was not off during the previous time step.

Hebbian model

Where applicable, parameter values were the same as for the drive-reinforcement model except that c=0.5, the initial synaptic weight values were 0.0, and there was no lower bound on the synaptic weights.

Sutton-Barto model

Where applicable, parameter values were the same as for the drive-reinforcement model except that c=0.5, α=0.9, the initial synaptic weight values were 0.0, and there was no lower bound on the synaptic weights.

Table 1.

Timing of the CS-US Configurations in Figures 4-20.

Figure Number	CS and US timing (time step of onset/time step of offset/trials during which stimulus was present)			
	CS_1	CS_2	CS_3	US
4(a)	10/15/1-50	----------	----------	14/15/1-50
4(b)	10/14/1-50	----------	----------	14/15/1-50
4(c)	10/14/1-120	----------	----------	14/15/1-120
5(a)	10/13/1-100	20/24/1-100	30/35/1-100	13/14/1-100
				23/24/1-100
				33/34/1-100
5(b)	10/13/1-300	20/24/1-300	30/35/1-300	13/14/1-300
				23/24/1-300
				33/34/1-300

(table continues)

Figure Number	CS and US timing (time step of onset/time step of offset/trials during which stimulus was present)			
	CS_1	CS_2	CS_3	US
5(c)	10/13/1-100	20/24/1-100	30/35/1-100	13/14/1-100
				23/24/1-100
				33/34/1-100
6	4/5/1-50	15/16/1-50	28/29/1-50	5/6/1-50
				16/19/1-50
				29/34/1-50
7	10/13/1-60	20/23/1-60	30/33/1-60	13/14/1-60
				23/24/1-60
				33/34/1-60
8	10/13/1-100	20/23/1-100	30/33/1-100	13/14/1-100
				23/24/1-100
				33/34/1-100

(table continues)

Figure Number	CS and US timing (time step of onset/time step of offset/trials during which stimulus was present)			
	CS_1	CS_2	CS_3	US
9	4/7/1-50	44/47/1-50	84/87/1-50	7/8/1-50
	14/17/1-50	54/57/1-50	94/97/1-50	17/18/1-50
	24/27/1-50	64/67/1-50	104/107/1-50	27/28/1-50
	34/37/1-50	74/77/1-50	114/117/1-50	37/38/1-50
				47/48/1-50
				67/68/1-50
				87/88/1-50
10	10/13/1-120	17/19/1-120	24/25/1-120	13/14/1-120
				20/21/1-120
				27/28/1-120

(table continues)

Figure Number	CS and US timing (time step of onset/time step of offset/trials during which stimulus was present)			
	CS_1	CS_2	CS_3	US
11	4/10/1-60	13/20/1-60	24/33/1-60*	4/10/1-60
				14/20/1-60
				27/33/1-60
				40/46/1-60
				57/63/1-60
12	10/15/1-200	7/12/61-200	----------	13/15/1-60
13	10/13/1-300	20/23/71-300	----------	13/16/1-200
	20/23/71-300			
14	10/13/1-200	----------	----------	13/16/1-70
				13/16/141-200
15(a)	10/14/1-25	----------	----------	12/16/1-25
15(b)	10/14/1-25	----------	----------	8/12/1-25

* CS_4: 35/46/1-60; CS_5: 51/63/1-60

(table continues)

Figure Number	CS and US timing (time step of onset/time step of offset/trials during which stimulus was present)			
	CS_1	CS_2	CS_3	US
16	10/13/1-160	10/13/101-160	----------	13/14/1-160
17	10/13/1-50	10/13/1-50	10/13/1-50	13/14/1-50
18(a)	10/12/1-80	20/22/1-80	----------	12/13/1-80
	20/22/1-80			22/23/1-80
18(b)	10/12/1-80	20/22/1-80	----------	22/23/1-80
	20/22/1-80			
19(a)	10/12/1-100	20/22/1-100	10/12/1-100	12/13/1-100
			20/22/1-100	

(table continues)

Figure Number	CS and US timing (time step of onset/time step of offset/trials during which stimulus was present)			
	CS_1	CS_2	CS_3	US
19(b)	5/7/1-20	15/17/1-20	5/7/1-20	7/8/1-20
	25/27/1-20	35/37/1-20	15/17/1-20	17/18/1-20
			25/27/1-20	
			35/37/1-20	
20	10/13/1-160	10/13/101-160	----------	13/14/1-160

END

FILMED

MARCH, 1988

DTIC